Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 23 (2021), Issue 2 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Original Papers

Letters to the Editor

Corrigenda and Addendas

Review

Machine Learning–Based Early Warning Systems for Clinical Deterioration: Systematic Scoping Review

Sankavi Muralitharan[1,2], MPharm, MSc; Walter Nelson[1*], BSc; Shuang Di[1,3*], BSc, MEd, MSc; Michael McGillion[4,5], BScN, PhD; PJ Devereaux[5,6], MD, PhD, FRCPC; Neil Grant Barr[7], BA, MSc, PhD; Jeremy Petch[1,5,8,9], HBA, MA, PhD

[1]Centre for Data Science and Digital Health, Hamilton Health Sciences, Hamilton, ON, Canada
[2]DeGroote School of Business, McMaster University, Hamilton, ON, Canada
[3]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[4]School of Nursing, McMaster University, Hamilton, ON, Canada
[5]Population Health Research Institute, Hamilton, ON, Canada
[6]Departments of Health Evidence and Impact and Medicine, McMaster University, Hamilton, ON, Canada
[7]Health Policy and Management, DeGroote School of Business, McMaster University, Hamilton, ON, Canada
[8]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[9]Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[*]these authors contributed equally

Corresponding Author:
Sankavi Muralitharan, MPharm, MSc
Centre for Data Science and Digital Health
Hamilton Health Sciences
293 Wellington St. N
Hamilton, ON, L8L 8E7
Canada
Phone: 1 2897882965
Email: sankavi_22@hotmail.com

Abstract

Background: Timely identification of patients at a high risk of clinical deterioration is key to prioritizing care, allocating resources effectively, and preventing adverse outcomes. Vital signs–based, aggregate-weighted early warning systems are commonly used to predict the risk of outcomes related to cardiorespiratory instability and sepsis, which are strong predictors of poor outcomes and mortality. Machine learning models, which can incorporate trends and capture relationships among parameters that aggregate-weighted models cannot, have recently been showing promising results.

Objective: This study aimed to identify, summarize, and evaluate the available research, current state of utility, and challenges with machine learning–based early warning systems using vital signs to predict the risk of physiological deterioration in acutely ill patients, across acute and ambulatory care settings.

Methods: PubMed, CINAHL, Cochrane Library, Web of Science, Embase, and Google Scholar were searched for peer-reviewed, original studies with keywords related to “vital signs,” “clinical deterioration,” and “machine learning.” Included studies used patient vital signs along with demographics and described a machine learning model for predicting an outcome in acute and ambulatory care settings. Data were extracted following PRISMA, TRIPOD, and Cochrane Collaboration guidelines.

Results: We identified 24 peer-reviewed studies from 417 articles for inclusion; 23 studies were retrospective, while 1 was prospective in nature. Care settings included general wards, intensive care units, emergency departments, step-down units, medical assessment units, postanesthetic wards, and home care. Machine learning models including logistic regression, tree-based methods, kernel-based methods, and neural networks were most commonly used to predict the risk of deterioration. The area under the curve for models ranged from 0.57 to 0.97.

Conclusions: In studies that compared performance, reported results suggest that machine learning–based early warning systems can achieve greater accuracy than aggregate-weighted early warning systems but several areas for further research were identified. While these models have the potential to provide clinical decision support, there is a need for standardized outcome measures to allow for rigorous evaluation of performance across models. Further research needs to address the interpretability of model outputs

by clinicians, clinical efficacy of these systems through prospective study design, and their potential impact in different clinical settings.

(J Med Internet Res 2021;23(2):e25187) doi:10.2196/25187

KEYWORDS

machine learning; early warning systems; clinical deterioration; ambulatory care; acute care; remote patient monitoring; vital signs; sepsis; cardiorespiratory instability; risk prediction

Introduction

Patient deterioration and adverse outcomes are often preceded by abnormal vital signs [1-3]. These warning signs frequently appear a few hours to a few days before the event, which can provide sufficient time for intervention. In response, clinical decision support early warning systems (EWS) have been developed that employ periodic observations of vital signs along with a predetermined criteria or cut-off range for alerting clinicians of patient deterioration [4].

EWS typically employ heart rate (HR), respiratory rate (RR), blood pressure (BP), peripheral oxygen saturation (SpO_2), temperature, and sometimes the level of consciousness [5]. Aggregate-weighted EWS incorporate several vital signs and other patient characteristics with clearly defined thresholds. Weights are assigned to each of these vital signs and characteristics based on a threshold, and an overall risk score is calculated by adding each of the weighted scores [6].

Some of the commonly used aggregate-weighted EWS for predicting cardiorespiratory insufficiency and mortality are the Modified Early Warning Score (MEWS) [7], National Early Warning Score (NEWS) [8], and Hamilton Early Warning Score [9], which all incorporate vital signs and the level of consciousness (Alert, Verbal, Pain, Unresponsive [AVPU]) but have varying thresholds for assigning scores.

The predictive ability of aggregate-weighted EWS has limitations. First, the scores indicate the present risk of the patient but do not incorporate trends nor provide information about the possible risk trajectory [6]; thus, the scores do not communicate whether the patient is improving or deteriorating and the rate of this change [10]. Second, these scores do not capture any correlations between the parameters, as the score for each parameter is calculated independently through simple addition [6] (eg, HR or RR can be interpreted differently when body temperature is taken into consideration).

A newer approach to EWS relies on machine learning (ML). ML models learn patterns and relationships directly from data rather than relying on a rule-based system [11]. Unlike aggregate-weighted EWS, ML models are computationally intensive, but can incorporate trends in risk scores, adjust for varying numbers of clinical covariates, and be optimized for different care settings and populations [12]. Like other EWS, ML models can be integrated into electronic health records to analyze vital sign measurements continuously and provide predictions of patient outcomes as part of a clinical decision support system [13].

Two systematic reviews in 2019 [14,15] evaluated the ability of ML models to predict clinical deterioration in adult patients using vital signs. The review by Brekke et al [15] examined the utility of trends within intermittent vital sign measurements from adult patients admitted to all hospital wards and emergency departments (ED) but identified only 2 retrospective studies that met their inclusion criteria. The review identified that vital sign trends were of value in detecting clinical deterioration but concluded that there is a lack of research in intermittently monitored vital sign trends and highlighted the need for controlled trials.

The review conducted by Linnen et al [14] compared the accuracy and workload of ML-based EWS with that of aggregate-weighted EWS. This review focused on studies that reported adult patient transfers to intensive care units (ICUs) or mortality as the outcome(s) and excluded all other clinical settings; 6 studies were identified that reported the performance metrics for both the ML-based EWS and aggregate-weighted EWS. The review identified that ML modelling consistently performed better than aggregate-weighted models while generating clinical workload. They also highlighted the need for standardized performance metrics and deterioration outcome definitions.

These are important findings, but to date no review has systematically reviewed the evidence from studies using ML-based EWS using vital sign measurements of varying frequencies, across different care settings and clinical outcomes in order to identify common methodological trends and limitations with current approaches to generate recommendations for future research in this area.

The objective of this study was to scope the state of research in ML-based EWS using vital signs data for predicting the risk of physiological deterioration in patients across acute and ambulatory care settings and to identify directions for future research in this area.

Methods

A systematic scoping review was conducted by following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for scoping reviews (PRISMA-ScR) framework [16]. This process provides an analysis of the available research, current state of utility of ML-based EWS, challenges facing their clinical implementation, and how they compare to aggregate-weighted EWS by identifying, synthesizing, and appraising the relevant evidence in the area. The literature search, assessment of eligibility of full-text articles, inclusion in the review, and extraction of study data were carried out by a single author.

Search Strategy

We searched PubMed, CINAHL, Cochrane Library, Web of Science, Embase, and Google Scholar for peer-reviewed studies without using any filters for study design and language. Searches were also conducted without any date restrictions. The reference lists of all studies that met the inclusion criteria were screened for additional articles. The search strategy involved a series of searches using a combination of relevant keywords and synonyms, including "vital signs," "clinical deterioration," and "machine learning." See Multimedia Appendix 1 for search terms.

Eligibility Criteria

The inclusion criteria covered the following:

- Peer-reviewed studies evaluating continuous or intermittent vital sign monitoring in adult patients so that all data collection or sampling frequencies (eg, 1 measurement per minute vs 1 measurement every 2 hours) wedre taken into consideration;
- Studies conducted using data gathered from all acute and ambulatory care settings including medical or surgical hospital wards, ICUs, step-down units, ED, and in-home care;
- Quantitative, observational, retrospective, and prospective cohort studies and randomized controlled trials;
- Studies that involved ML or multivariable statistical or ML models and reported some model performance measure (eg, area under the curve) [17];
- Studies that reported mortality or any outcomes related to clinical deterioration so that EWS models and performance can be examined for all explored outcomes.

The exclusion criteria included the following:

- Studies that used any laboratory values as predictors for the ML-based EWS, as this review focuses on examining time-sensitive predictions of clinical deterioration using patient parameters that are readily available across all care settings;
- Studies involving pediatric or obstetric populations due to these patients having different or altered physiologies that cannot be compared to standard adult patients;
- Qualitative studies, reviews, preprints, case reports, commentaries, or conference proceedings.

Study Selection

References from the preliminary searches were handled using Mendeley reference management software. After duplicates were removed, titles and abstracts were screened to assess preliminary eligibility. Eligible studies were then read in full length to be assessed against the inclusion and exclusion criteria.

Data Extraction

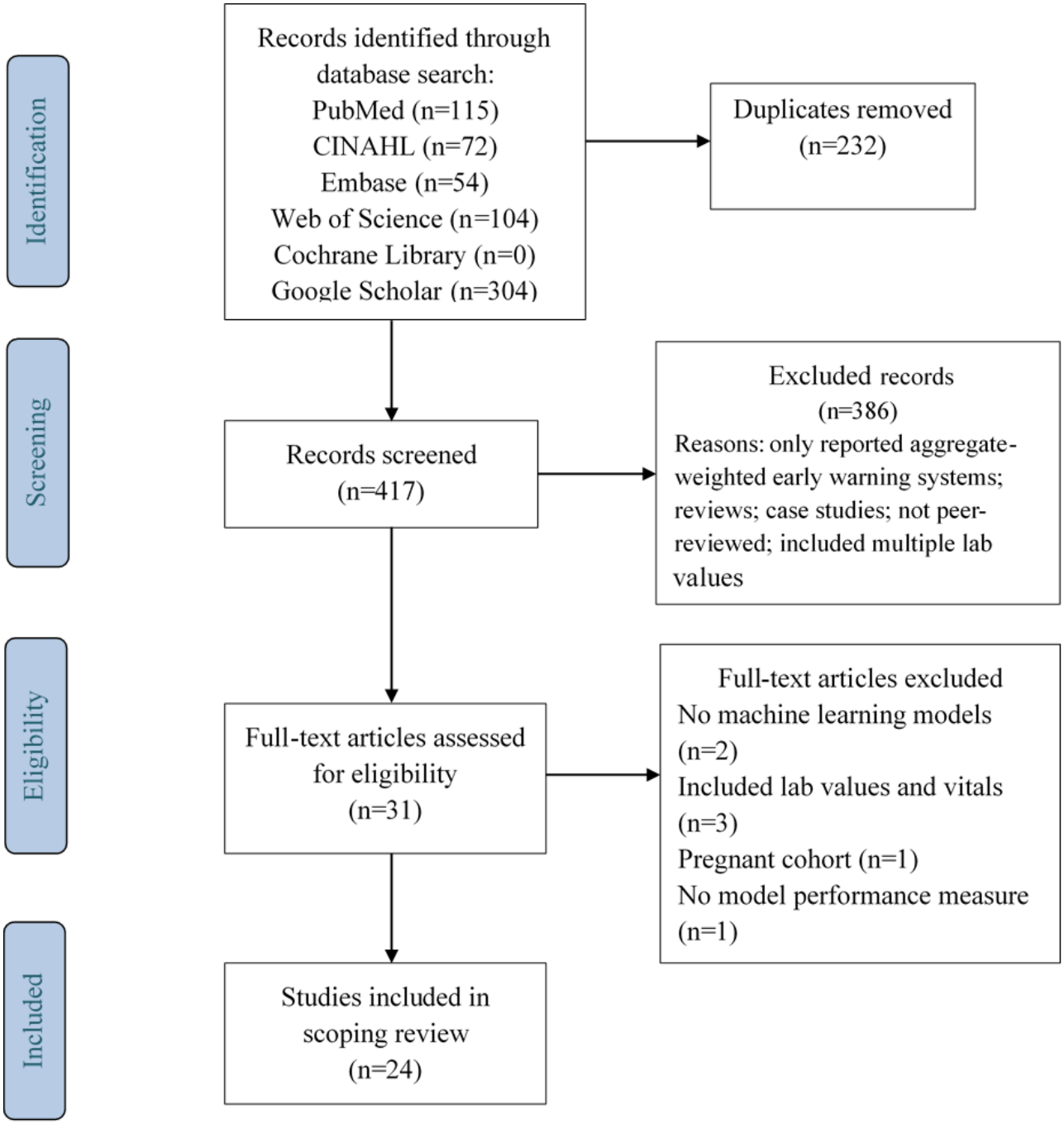
Data were extracted from eligible studies using an extraction sheet that followed the PRISMA [18] and Cochrane Collaboration guidelines for systematic reviews [19] and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines [20] for the reporting of predictive models. Study characteristics, setting, demographics, patient outcomes, ML model characteristics, and model performance data were extracted. The model performance results were extracted from the validation data set rather than from the model derivation or training data set to decrease the potential for model overfitting. When studies explored multiple ML models, the model with the best performance was selected for reporting and comparison. If studies compared the performance of ML models to aggregate-weighted EWS, then the performance data of these warning systems were also extracted.

Results

Search Results and Study Selection

The search for "vital signs" AND "clinical deterioration" AND "machine learning" using the same query terms and filters identified 417 studies after duplicate removal. During the title and abstract screening process, 386 studies were excluded. Of the 31 full-text articles that were assessed, 7 studies were excluded for not meeting the eligibility criteria: 2 studies did not use ML models to predict deterioration, 3 studies included vital sign measurements in addition to laboratory values as predictors, 1 study focused on a cohort of pregnant women, and 1 study did not meet the criteria for model performance measures. A review of the reference lists of the 24 selected studies did not yield any additional studies fulfilling the eligibility criteria (refer to Figure 1).

Figure 1. PRISMA flowchart of the search strategy and study selection.



Study Characteristics

Of the selected studies, 23 conducted a retrospective analysis of the vital signs data, while 1 study [21] used a prospective cohort study design. Seventeen studies only analyzed continuous vital signs measurements collected through wearable devices and bedside monitors, whereas 3 [22-24] studies analyzed vital signs that were collected both manually and intermittently by clinical staff. Two studies [25,26] analyzed vital signs that were collected both continuously and intermittently, while the remaining 2 studies did not report how the vital sign data were collected.

Studies were conducted in a variety of settings within hospitals while the study by Larburu et al [22] was conducted in an ambulatory setting. While 3 studies [27-29] aimed to develop a remote home-based monitoring tool, the vital sign data used were obtained from the Medical Information Mart for Intensive Care (MIMIC and MIMIC-II) databases [30,31] consisting of data captured from patient monitors in different ICUs. Regarding location, 5 studies [24,26,32-34] were conducted on general wards, 4 studies [11,23,35,36] were conducted in EDs, 7 studies [26,34,37-41] were conducted in ICUs, 2 studies [25,42] were conducted in postoperative wards, and 4 studies [21,43-45] in acute stay wards (medical admission unit, step-down units). Cohort sizes for the studies ranged from 12 patients [39] to 10,967,518 patient visits [11] (refer to Table 1).

Table 1. Study characteristics.

Authors, year	Setting(s)	Data collection	Cohort description	Event rate	Study purpose	Predictors	Measurement frequency	Outcome
Badriyah et al, 2014 [45]	Medical assessment unit for 24 hours	Personal digital assistants running VitalPAC software	35,585 admissions	199 (0.56%), cardiac arrest; 1161 (3.26%) unanticipated ICU[a] admissions; 1789 (5.02%) deaths; 3149 (8.85%) any outcome	Compare the performance of a decision tree analysis with NEWS[b]	HR[c], RR[d], SBP[e], temperature, SpO_2, AVPU[f] level, % breathing air at the time of SpO_2 measurement	Not specified	Cardiac arrest, unanticipated ICU admission, or death, each within 24 hours of a given vital sign observation
Chen at al, 2017 [44]	Step-down unit	Bedside monitors	1880 patients (1971 admissions)	997 patients (53%) or 1056 admissions (53.6%) who experienced CRI[g] events	Describe the dynamic and personal character of CRI risk evolution observed through continuous vital sign monitoring of individual patients	HR, RR, SPO_2 (at 1/20 Hz), SBP, DBP[h]	Every 2 hours	CRI
Churpek et al, 2016 [24]	All wards at the University of Chicago and 4 North Shore University Health System hospitals	Data collected manually, documented electronically	269,999 admissions	16,452 outcomes (6.09%)	Whether adding trends improves accuracy of early detection of clinical deterioration and which methods are optimal for modelling trends	Temperature, HR, RR, SpO_2, DBP, SBP	Every 4 hours	Development of critical illness on the wards: deaths, cardiac arrest, ICU transfers
Chiew et al, 2019 [23]	ED[i] at Singapore general hospital	Measurements at triage; hospital EHR[j]	214 patients	40 patients (18.7%) met outcome	Compare the performance of HR variability–based machine learning models vs conventional risk stratification tools to predict 30-day mortality	Age, gender, ethnicity, temperature, HR, RR, SBP, DBP, GCS[k], HR variability	At triage	30-day mortality due to sepsis
Chiu et al, 2019 [42]	Postoperative surgical wards at 4 UK adult cardiac surgical centers	VitalPac to electronically capture patients' vital signs	Adults undergoing risk-stratified major cardiac surgery, n=13,631	578 patients (4.2%) with an outcome; 499 patients (3.66%) with unplanned ICU readmissions	Using logistic regression to model the association of NEWS variables with a serious patient event in the subsequent 24 hours; secondary objectives: comparing the discriminatory power of each model for events in the next 6 hours or 12 hours	RR, SpO_2, SBP, HR, temperature, consciousness level	Not specified	Death, cardiac arrest, unplanned ICU readmissions
Clifton et al, 2014 [25]	Postoperative ward of the cancer center, Oxford University Hospitals NHS[l] Trust, United Kingdom	Continuous vitals monitored by wearable devices; intermittent vitals monitored manually by ward staff	200 patients in the postoperative ward following upper gastrointestinal cancer surgery	Not specified	Using continuous vitals monitoring to provide early warning of physiological deterioration, such that preventative clinical action may be taken	SpO_2, HR (256 Hz), BP, RR	Continuously (SpO_2, HR), intermittently (BP, RR)	Physiological deterioration

Authors, year	Setting(s)	Data collection	Cohort description	Event rate	Study purpose	Predictors	Measurement frequency	Outcome
Desautels et al, 2016 [37]	Beth Israel Deaconess Medical Center ICU	ICU bedside monitors and medical records (MIMIC^m-III)	22,853 ICU stays	2577 (11.28%) stays with confirmed sepsis	Validate a sepsis prediction method, InSight, for the new Sepsis-3 definitions and make predictions using a minimal set of variables	GCS, HR, RR, SpO_2, temperature, invasive and noninvasive SBP and DBP	At least 1 measurement per hour	Onset of sepsis
Forkan et al, 2017 [28]	Beth Israel Deaconess Medical Center ICU	ICU bedside monitors and medical records (MIMIC-II)	1023 patients	Not specified	Develop a probabilistic model for predicting the future clinical episodes of a patient using observed vital sign values prior to the clinical event	HR, SBP, DBP, mean BP, RR, SpO_2	All samples converted to per-minute sampling	Abnormal clinical events
Forkan et al, 2017 [27]	Beth Israel Deaconess Medical Center ICU	ICU bedside monitors and medical records (MIMIC & MIMIC-II)	85 patients	Not specified	Develop an intelligent method for personalized monitoring and clinical decision support through early estimation of patient-specific vital sign values	HR, SBP, DBP, mean BP, RR, SpO_2	Per-minute sampling	Patient-specific anomalies, disease symptoms, and emergencies
Forkan et al, 2017 [29]	Beth Israel Deaconess Medical Center ICU	ICU bedside monitors and medical records (MIMIC-II)	4893 patients	Not specified	Build a prognostic model, ViSiBiD, that can accurately identify dangerous clinical events of a home-monitored patient in advance	HR, SBP, DBP, mean BP, RR, SpO_2	Per-minute sampling	Dangerous clinical events
Guillame-Bert et al, 2017 [43]	Step-down unit	Bedside monitor measurements over 8 weeks	297 admissions	127 patients (43%) exhibited at least 1 real event during their stay	Forecast CRI utilizing data from continuous monitoring of physiologic vital sign measurements	HR, RR, SPO_2, SBP, DBP, mean BP	Every 20 seconds (HR, RR, SPO_2), every 2 hours (SBP, DBP, and mean BP)	At least 1 event threshold limit criteria exceeded for >80% of last 3 minutes
Ho et al, 2017 [38]	Beth Israel Deaconess Medical Center ICU	ICU bedside monitors and medical records (MIMIC-II)	763 patients	197 patients (25.8%) experienced a cardiac arrest event	Build a cardiac arrest risk prediction model capable of early notification at time z (z ≥5 hours prior to the event)	Temperature, SpO_2, HR, RR, DBP, SBP, pulse pressure index	1 reading per hour	Cardiac arrest
Jang et al, 2019 [35]	ED visits to a tertiary academic hospital	EHR data from ED visits	Nontraumatic ED visits	374,605 eligible ED visits of 233,763 patients; 1097 (0.3%) patients with cardiac arrest	Develop and test artificial neural network classifiers for early detection of patients at risk of cardiac arrest in EDs	Age, sex, chief complaint, SBP, DBP, HR, RR, temperature, AVPU	Not specified	Development of cardiac arrest within 24 hours after prediction

Authors, year	Setting(s)	Data collection	Cohort description	Event rate	Study purpose	Predictors	Measurement frequency	Outcome
Kwon et al, 2018 [26]	Cardiovascular teaching hospital and community general hospital	Data collected manually by staff on general wards, by bedside monitors in ICUs	52,131 patients	419 patients (0.8%) with cardiac arrest; 814 (1.56%) deaths without attempted resuscitation	Predict whether an input vector belonged within the prediction time window (0.5-24 hours before the outcome)	SBP, HR, RR, temperature	3 times a day on general wards, every 10 minutes in ICUs	Primary outcome: first cardiac arrest; secondary outcome: death without attempted resuscitation
Kwon et al, 2018 [11]	151 EDs in Korea	Korean National Emergency Department Information System (NEDIS)	10,967,518 ED visits	153,217 (1.4%) in-hospital deaths; 625,117 (5.7%) critical care admissions; 2,964,367 (27.0%) hospitalizations	Validate that a DTAS[n] identifies high-risk patients more accurately than existing triage and acuity scores	Age, sex, chief complaint, time from symptom onset to ED visit, arrival mode, trauma, initial vital signs (SBP, DBP, HR, RR, temperature), mental status	At ED admission	Primary outcome: in-hospital mortality; secondary outcome: critical care; tertiary outcome: hospitalization
Larburu et al, 2018 [22]	OSI Bilbao-Basurto (Osakidetza) Hospital and ED admissions, ambulatory	Collected manually by clinicians and patients	242 patients	202 predictable decompensations	Prevent mobile heart failure patients' decompensation using predictive models	SBP, DBP, HR, SaO_2, weight	At diagnosis and 3-7 times per week in ambulatory patients	Heart failure decompensation
Li et al, 2016 [39]	Beth Israel Deaconess Medical Center ICU	ICU bedside monitors and medical records (MIMIC-II)	12 patients	Not specified	Adaptive online monitoring of patients in ICUs	HR, SBP, DBP, MAP[o], RR	At least 1 measurement per hour	Signs of deterioration
Liu et al, 2014 [36]	ED of a tertiary hospital in Singapore	Manual vital measurements by nurses or physicians	702 patients with undifferentiated, nontraumatic chest pain	29 (4.13%) patients met primary outcome	Discover the most relevant variables for risk prediction of major adverse cardiac events using clinical signs and HR variability	SBP, RR, HR	Not specified	Composite of events such as death and cardiac arrest within 72 hours of arrival at the ED
Mao et al, 2018 [34]	ICU, inpatient wards, outpatient visits	UCSF[p] dataset:inpatient and outpatient visits; MIMIC-III: ICU bedside monitors	UCSF: 90,353 patients; MIMIC-III: 21,604 patients	UCSF: 1179 (1.3%) sepsis, 349 (0.39%) severe sepsis, 614 (0.68%) septic shock; MIMIC-III: sepsis (1.91%), severe sepsis (2.82%), septic shock (4.36%)	Sepsis prediction	SBP, DBP, HR, RR, SpO_2, temperature	Hourly	Sepsis, severe sepsis, septic shock
Olsen et al, 2018 [46]	PACU[q], Rigshospitalet, University of Copenhagen, Denmark	IntelliVue MP5, BM-EYE Nexfin bedside monitors during admission to post anesthetic care unit	178 patients	160 (89.9%) had ≥1 microevent occurring during admission; 116 patients (65.2%) had ≥1 microevent with a duration >15 minutes	Develop a predictive algorithm detecting early signs of deterioration in the PACU using continuously collected cardiopulmonary vital signs	SpO_2, SBP, HR, MAP	Every minute (SpO_2, SBP, HR), every 15 minutes (MAP)	Signs of deterioration

Authors, year	Setting(s)	Data collection	Cohort description	Event rate	Study purpose	Predictors	Measurement frequency	Outcome
Shashikumar et al, 2017 [40]	Adult ICU units	ICU bedside monitors, Bedmaster system; up to 24 hours of monitoring	Patients with unselected mixed surgical procedures	242 sepsis cases	Predict onset of sepsis 4 hours ahead of time, using commonly measured vital signs	MAP, HR, SpO_2, SBP, DBP, RR, GCS, temperature, comorbidity, clinical context, admission unit, surgical specialty, wound type, age, gender, weight, race	≥1 measurement per hour	Onset of sepsis
Tarassenko et al, 2006 [32]	General wards at John Radcliffe Hospital in Oxford, United Kingdom	Bedside monitors for at least 24 hours per patient	150 general-ward patients	Not specified	A real-time automated system, BioSign, which tracks patient status by combining information from vital signs	HR, RR, SpO_2, skin temperature, average SBP - average DBP	Every 30 minutes (BP), every 5 seconds (other vitals)	Signs of deterioration
Van Wyk et al, 2017 [33]	Methodist LeBonheur Hospital, Memphis, TN	Bedside monitors: Cerner CareAware iBus system	2995 patients	343 patients (11.5%) diagnosed with sepsis	Classify patients into sepsis and nonsepsis groups using data collected at various frequencies from the first 12 hours after admission	HR, MAP, DBP, SBP, SpO_2, age, race, gender, fraction of inspired oxygen	Every minute	Sepsis detection
Yoon et al, 2019 [41]	Beth Israel Deaconess Medical Center ICU	ICU bedside monitors and medical records (MIMIC-II)	2809 subjects	787 tachycardia episodes	Predicting tachycardia as a surrogate for instability	Arterial DBP, arterial SBP, HR, RR, SpO_2, MAP	1/60 Hz or 1 Hz	Tachycardia episode

[a]ICU: intensive care unit.
[b]NEWS: National Early Warning Score.
[c]HR: heart rate.
[d]RR: respiratory rate.
[e]SBP: systolic blood pressure.
[f]AVPU: alert, verbal, pain, unresponsive.
[g]CRI: cardiorespiratory instability.
[h]DBP: diastolic blood pressure.
[i]ED: emergency department.
[j]EHR: electronic health record.
[k]GCS: Glasgow Coma Score.
[l]NHS: National Health Service.
[m]MIMIC: Medical Information Mart for Intensive Care.
[n]DTAS: Deep learning–based Triage and Acuity Score.
[o]MAP: mean arterial pressure.
[p]UCSF: University of California, San Francisco.
[q]PACU: postanesthesia care unit.

Predictor Variables

The most commonly used vital sign predictors were HR, RR, systolic BP, diastolic BP, SpO_2, body temperature, level of consciousness through either the Glasgow Coma Score or the AVPU scale, and mean arterial pressure. Measurement frequencies for these variables ranged from once every 5 seconds [32] in hospital wards to 3-7 times per week [22] in an ambulatory setting. Other commonly used predictors included age, gender, weight, ethnicity, chief complaint, and comorbidities.

Outcomes

The outcomes being predicted in most studies focused on cardiorespiratory insufficiency–related events. Cardiac arrest was the primary outcome in 7 [24,26,35,36,38,42,45] studies, while general cardiorespiratory deterioration or decompensation was the primary outcome in 5 studies [25,39,41,43,44]. Another commonly predicted outcome was sepsis, which included the time of onset of sepsis [34,37,40], severe sepsis [33,34], septic shock [34], and sepsis-related mortality [23]. Other outcomes explored within the studies include unanticipated ICU admissions [24,42,45], development of critical illness [24], general physiological deterioration [25,32,39,46], abnormal or dangerous clinical events [27-29], and mortality [11,24,42].

Outcomes were first identified, and baseline models were created using predefined parameter thresholds (ground truth) consistent with the MEWS [23,26,35] or NEWS [23,42,46] criteria for cardiorespiratory instability and general physiological deterioration, while the sepsis-related outcomes were identified based on the thresholds set within the systemic inflammatory response syndrome [34], quick Sequential Organ Failure Assessment (qSOFA) [23], and SOFA [37] criteria. Some studies [22,27-29,43,44] also used thresholds and criteria based on the population served by their individual care setting.

ML Models and Performance

All included studies consider the prediction of deterioration risk to be a classification task and therefore use different types of classification models in the process, including tree-based models, linear models, kernel-based methods, and neural networks (refer to Table 2 for a full inventory of methods used, model performance achieved, and prediction windows, and see Multimedia Appendix 2 for a description of ML methods).

Measures used to assess model performance varied across the studies. The most common measure was the area under the receiver operator characteristic (AUROC) along with model accuracy, sensitivity, and specificity. Area under the precision-recall, F-score, Hamming's score, and precision (positive predictive value) were reported less commonly.

Prediction windows ranged from 30 minutes to 30 days before an event.

Model performance varied substantially based on outcome measure being predicted (eg, cardiorespiratory insufficiency vs sepsis), ML method used (eg, linear vs tree-based), and prediction window (eg, 30 minutes before an event vs 4 hours before).

Table 2. Machine learning (ML) models and comparisons used for outcome prediction.

Study	Cohort	Event rate	ML model(s)	Missing data handling	Best ML model performance	ML model comparisons	Prediction window	Aggregate weighted EWS[a] comparisons
Badriyah et al, 2014 [45]	35,585 admissions	199 (0.56%), cardiac arrest; 1161 (3.26%) unanticipated ICU[b] admissions; 1789 (5.02%) deaths; 3149 (8.85%) any outcome	Decision tree analysis	Not specified	Decision tree predicted cardiac arrest: AUROC[c]=0.708; unanticipated ICU admission: AUROC=0.862; death: AUROC=0.899; any outcomes: AUROC=0.877	Not specified	Within 24 hours preceding events	NEWS[d] AUROC: cardiac arrest, 0.722; unanticipated ICU admission, 0.857; death, 0.894; any outcomes, 0.873
Chen at al, 2017 [44]	1880 patients (1971 admissions)	997 patients (53%) or 1056 admissions (53.6%) who experienced CRI[e] events	Variant of the random forest classification model using nonrandom splits	Not specified	Random forest AUC[f] initially remained constant (0.58-0.60), followed by an increasing trend, with AUCs rising from 0.57 to 0.89 during the 4 hours immediately preceding events	Logistic regression: AUC=0.7; lasso logistic regression: AUC=0.82	Within 4 hours preceding events	No comparison
Churpek et al, 2016 [24]	269,999 admissions	16,452 outcomes (6.09%)	Univariate analysis, bivariate analysis	Forward imputation, median value imputation	Trends increased model accuracy compared to a model containing only current vital signs (AUC 0.78 vs 0.74); vital sign slope improved AUC by 0.013	Not specified	Within 4 hours preceding events	No comparison
Chiew et al, 2019 [23]	214 patients	40 patients (18.7%) met outcome	K-nearest neighbor, random forest, adaptive boosting, gradient boosting, support vector machine	Not specified	Gradient boosting predicted 30-day sepsis-related mortality: F1 score=0.50, AUPRC=0.35, precision (PPV[g])=0.62, recall=0.5	K-nearest neighbor: F1 score=0.10, AUPRC=0.10, precision (PPV)=0.33, recall=0.6; random forest: F1 score=0.35, AUPRC=0.27, precision (PPV)=0.26, recall=0.56; adaptive boosting: F1 score=0.40, AUPRC=0.31, precision (PPV)=0.43, recall=0.38; SVM[h]: F1 score=0.43, AUPRC=0.29, precision (PPV)=0.33, recall=0.63	Within 30 days preceding event	SEDS[i]: F1=0.40, AUPRC=0.22; qSOFA[j]: F1=0.32, AUPRC=0.21; NEWS; F1=0.38, AUPRC=0.28; MEWS[k]: F1=0.30, AUPRC=0.25

Study	Cohort	Event rate	ML model(s)	Missing data handling	Best ML model performance	ML model comparisons	Prediction window	Aggregate weighted EWS[a] comparisons
Chiu et al, 2019 [42]	Adults undergoing risk-stratified major cardiac surgery (n=13,631)	578 patients (4.2%) with an outcome; 499 patients (3.66%) with unplanned ICU readmissions	Logistic regression	Observations with missing values were excluded	Logistic regression predicted the event 24 hours in advance: AUROC=0.779; 12 hours in advance: AUROC=0.815; 6 hours in advance: AUROC=0.841	Not specified	Within 24, 12, and 6 hours preceding event	NEWS: 24 hours before event, AUROC=0.754; 12 hours before event, AUROC=0.789; 6 hours before event, AUROC=0.813
Clifton et al, 2014 [25]	200 patients in the postoperative ward following upper gastrointestinal cancer surgery	Not specified	Classifiers, Gaussian process, one-class support vector machine, kernel estimate	Missing channels replaced by mean of that channel	SVM predicted deterioration: accuracy=0.94, partial AUC=0.28, sensitivity=0.96, specificity=0.93	Conventional SVM: accuracy=0.90, partial AUC=0.26, sensitivity=0.92, specificity=0.87; Gaussian mixture models: accuracy=0.9, partial AUC=0.24, sensitivity=0.97, specificity=0.84; Gaussian processes: accuracy=0.90, partial AUC=0.26, sensitivity=0.91, specificity=0.89; kernel density estimate: accuracy=0.91, partial AUC=0.26, sensitivity=0.94, specificity=0.87	Not specified	No comparison
Desautels et al, 2016 [37]	22,853 ICU stays	2577 (11.28%) stays with confirmed sepsis	Insight classifier	Carry forward imputation	Classifier predicts sepsis at onset: AUROC=0.880, APR[l]=0.6, accuracy=0.8; classifier predicts sepsis 4 hours before onset: AUROC=0.74, APR=0.28, accuracy=0.57	Not specified	Within 4 hours preceding event and at time of event onset	SIRS[m]: AUROC= 0.609, APR= 0.160; qSOFA: AUROC= 0.772, APR=0.277; MEWS: AUROC=0.803, APR=0.327; SAPS[n] II: AUROC=0.700, APR=0.225; SOFA: AUROC=0.725, APR=0.284

Study	Cohort	Event rate	ML model(s)	Missing data handling	Best ML model performance	ML model comparisons	Prediction window	Aggregate weighted EWS[a] comparisons
Forkan et al, 2017 [28]	1023 patients	Not specified	PCA[o] used to separate patients into multiple categories; hidden Markov Model adopted for probabilistic classification and future prediction	Data with consecutive missing values over a long period are eliminated	Hidden Markov Model event prediction: accuracy=97.8%, precision=92.3, sensitivity=97.7, specificity=98, F-score=95%	Neural network: accuracy=93%	Within 30 minutes preceding event	No comparison
Forkan et al, 2017 [27]	85 patients	Not specified	Multilabel classification algorithms are applied in classifier design; result analysis with J48 decision tree, random tree and sequential minimal optimization (SMO, a simplified version of SVM)	Where ≥1 vital signs data are missing while clean values of others are available, considered as recoverable and imputed using median-pass and k-nearest neighbor filter	Predictions across 24 classifier combinations yielded a Hamming score of 90%-95%; F1-micro average of 70.1%-84%; accuracy of 60.5%-77.7%	Not specified	Within 1 hour preceding event	No comparison
Forkan et al, 2017 [29]	4893 patients	Not specified	J48 decision tree, random forest, sequential minimal optimization, MapReduce random forest	Data with consecutive missing values over a long period are eliminated	Event prediction by random forest: within a 60-minute forecast horizon, F score=0.96, accuracy=95.86; within a 90-minute forecast horizon, F-score=0.95, accuracy=95.35; within a 120-minute forecast horizon, F-score=0.95, accuracy=95.18	J48 decision tree: within a 60-minute forecast horizon, F score=0.93, accuracy=92.46; within a 90-minute forecast horizon, F score=0.92, accuracy=91.59; within a 120-minute forecast horizon, F score=0.91, accuracy=91.30; Event prediction with sequential minimal optimization: within a 60-minute forecast horizon, F score=0.91, accuracy=90.72; within a 90-minute forecast horizon, F score=0.90, accuracy=90.08; within a 120-minute forecast horizon, F score=0.89, accuracy=89.23	1 hour preceding event	No comparison

Study	Cohort	Event rate	ML model(s)	Missing data handling	Best ML model performance	ML model comparisons	Prediction window	Aggregate weighted EWS[a] comparisons
Guillame-Bert et al, 2017 [43]	297 admissions	127 patients (43%) exhibited at least 1 real CRI event during their stay in the step-down unit	TITA[p] rules, rule fusion algorithm; mapping function from rule-based features to forecast model learned using random forest classifier	Not specified	Event forecast alert within 17 minutes, 51 seconds before onset of CRI (false alert every 12 hours); event forecast alert within 10 minutes, 58 seconds before onset of CRI (false alert every 24 hours)	Random forest: event forecast alert within 11 minutes, 25 seconds before onset of CRI (false alert every 12 hours); event forecast alert within 5 minutes, 52 seconds before onset of CRI (false alert every 24 hours)	Within 17 minutes, 51 seconds preceding CRI onset	No comparison
Ho et al, 2017 [38]	763 patients	197 patients (25.8%) experienced a cardiac arrest event	Temporal transfer learning-based model (TTL-Reg)	Imputed values based on the median from patients of the same gender and similar ages	TTL-Reg predicts events with an AUC of 0.63	Not specified	Within 6 hours preceding event	No comparison
Jang et al, 2019 [35]	Non-traumatic ED visits	374,605 eligible ED visits of 233,763 patients; 1097 (0.3%) patients with cardiac arrest	ANN[q] with multilayer perceptron, ANN with LSTM[r], hybrid ANN; comparison with random forest and logistic regression	Not specified	Event prediction: ANN with multilayer perceptron, AUROC=0.929; ANN with LSTM, AUROC=0.933; hybrid ANN, AUROC=0.936	Random forest, AUROC=0.923; logistic regression, AUROC=0.914	Within 24 hours preceding event	MEWS: AUROC=0.886
Kwon et al, 2018 [26]	52,131 patients	419 patients (0.8%) with cardiac arrest; 814 (1.56%) deaths without attempted resuscitation	3 RNN[s] layers with LSTM to deal with time series data; compared to random forest and logistic regression	Most recent value was used; if no value available, then median value used	Event prediction: RNNs, AUROC=0.85, AUPRC[t]=0.044	Random forest, AUROC=0.78, AUPRC=0.014; logistic regression, AUROC=0.613, AUPRC=0.007	30 minutes to 24 hours preceding event	MEWS: AUROC=0.603, AUPRC=0.003
Kwon et al, 2018 [11]	10,967,518 ED visits	153,217 (1.4%) in-hospital deaths; 625,117 (5.7%) critical care admissions; 2,964,367 (27.0%) hospitalizations	DTAS[u] using multilayer perceptron with 5 hidden layers	Excluded	Event prediction: DTAS using multilayer perceptron, AUROC=0.935, AUPRC=0.264	Random forest: AUROC= 0.89, AUPRC= 0.14; logistic regression: AUROC= 0.89, AUPRC=0.16	Not specified	Korean triage and acuity score: AUROC =0.785, AUPRC=0.192; MEWS: AUROC=0.810, AUPRC=0.116;
Larburu et al, 2018 [22]	242 patients	202 predictable decompensations	Naïve Bayes, decision tree, random forest, SVM	Not specified	Decompensation event prediction: naïve Bayes, AUC=67%	Decision tree, neural network, random forest, support vector machine, stochastic gradient descent	Not specified	No comparison

Study	Cohort	Event rate	ML model(s)	Missing data handling	Best ML model performance	ML model comparisons	Prediction window	Aggregate weighted EWS[a] comparisons
Li et al, 2016 [39]	12 patients	Not specified	L-PCA (combination of just-in-time learning and PCA)	Not specified	Fault detection rate with L-PCA: 20% higher than with PCA; 47% higher than with fast moving-window PCA; best detection rate achieved was 99.8%	Not specified	Not specified	No comparison
Liu et al, 2014 [36]	702 patients with undifferentiated, non-traumatic chest pain	29 (4.13%) patients met primary outcome	Novel variable selection framework based on ensemble learning; random forest was the independent variable selector for creating the decision ensemble	Not specified	Event prediction with ensemble learning model: AUC=0.812, cutoff score=43, sensitivity=82.8%, specificity=63.4%	Not specified	Within 72 hours of arrival at ED	TIMI[v]: AUC=0.637; MEWS: AUC=0.622
Mao et al, 2018 [34]	UCSF[w]: 90,353 patients; MIMIC[x]-III: 21,604 patients	UCSF: 1179 (1.3%) sepsis, 349 (0.39%) severe sepsis, 614 (0.68%) septic shock; MIMIC-III: sepsis (1.91%), severe sepsis (2.82%), septic shock (4.36%)	Gradient tree boosting + transfer learning using MIMIC-III as source and UCSF as target	Carry forward imputation	Detection with gradient tree boosting: AUROC=0.92 for sepsis; AUROC=0.87 for severe sepsis at onset; AUROC=0.96 for septic shock 4 hours before; AUROC=0.85 for severe sepsis prediction 4 hours before	Not specified	At onset of sepsis and severe sepsis; within 4 hours preceding septic shock and severe sepsis	MEWS: AUROC=0.76; SOFA: AUROC=0.65; SIRS: AUROC=0.72
Olsen et al, 2018 [46]	178 patients	160 (89.9%) had ≥1 microevent occurring during admission; 116 patients (65.2%) had ≥1 microevent with a duration >15 minutes	Random forest classifier	Not specified	Detection of early signs of deterioration with random forest: accuracy=92.2%, sensitivity=90.6%, specificity=93.0%, AUROC=96.9%	Not specified	Not specified	Compared with hospital's current alarm system: number of false alarms decreased by 85%, number of missed early signs of deterioration decreased by 73%

Study	Cohort	Event rate	ML model(s)	Missing data handling	Best ML model performance	ML model comparisons	Prediction window	Aggregate weighted EWS[a] comparisons
Shashikumar et al, 2017 [40]	Patients with unselected mixed surgical procedures	242 sepsis cases	Elastic net logistic classifier	Median values (if multiple measurement were available); otherwise, the old values were kept (sample-and-hold interpolation); mean imputation for replacing all remaining missing values	Event prediction: elastic net logistic classifier using entropy features alone, AUROC=0.67, accuracy=47%; elastic net logistic classifier using social demographics + EMR[y] features, AUROC=0.7, accuracy=50%; elastic net logistic classifier using all features, AUROC=0.78, accuracy=61%	Not specified	4 hours prior to onset	No comparison
Tarassenko et al, 2006 [32]	150 general-ward patients	Not specified	Biosign; data fusion method: probabilistic model of normality in five dimensions	Historic, median filtering	95% of Biosign alerts were classified as “True” by clinical experts	Not specified	Within 120 minutes of event	No comparison
Van Wyk et al, 2017 [33]	2995 patients	343 patients (11.5%) diagnosed with sepsis	CNN[z] (constructed images using raw patient data) with random dropout to reduce overfitting; multilayer perceptron with random dropout between layers to avoid overfitting	Not specified	Event classification with a 1-minute observation frequency: CNN, accuracy=86.1%; event classification with a 10-minute observation frequency: CNN, accuracy=78.2%	Event classification with a 1-minute observation frequency: multilayer perceptron, accuracy=76.5%; event classification with a 10-minute observation frequency: multilayer perceptron, accuracy=71%	Not specified	No comparison

Study	Cohort	Event rate	ML model(s)	Missing data handling	Best ML model performance	ML model comparisons	Prediction window	Aggregate weighted EWS[a] comparisons
Yoon et al, 2019 [41]	2809 subjects	787 tachycardia episodes	Regularized logistic regression and random forest classifiers	Discrete Fourier transform, cubic-spline interpolation of heart rate and respiratory rate data for missing data as long as ≥20% of the data were available	Event prediction: random forest, AUC=0.869, accuracy=0.806	Logistic regression with L1 regularization, AUC=0.8284, accuracy=0.7668	Within 3 hours preceding onset	No comparison

[a]EWS: early warning system.
[b]ICU: intensive care unit.
[c]AUROC: area under the receiver operator characteristic.
[d]NEWS: National Early Warning Score.
[e]CRI: cardiorespiratory instability.
[f]AUC: area under the curve.
[g]PPV: positive predictive value.
[h]SVM: support vector machine.
[i]SEDS: Singapore Emergency Department Sepsis.
[j]qSOFA: quick Sequential Organ Failure Assessment.
[k]MEWS: Modified Early Warning Score.
[l]APR: area under the precision-recall curve.
[m]SIRS: systemic inflammatory response syndrome.
[n]SAPS II: simplified acute physiology score.
[o]PCA: principal component analysis.
[p]TITA: temporal interval tree association.
[q]ANN: artificial neural network.
[r]LSTM: long short-term memory.
[s]RNN: recurrent neural network.
[t]AUPRC: area under the precision-recall curve.
[u]DTAS: Deep learning–based Triage and Acuity Score.
[v]TIMI: Thrombolysis in Myocardial Infarction.
[w]UCSF: University of California, San Francisco.
[x]MIMIC: Medical Information Mart for Intensive Care.
[y]EMR: electronic medical record.
[z]CNN: convolutional neural network.

Comparison With Aggregate-Weighted EWS

Nine studies compared the performance of ML-based EWS with aggregate-weighted EWS. Studies exploring cardiorespiratory outcomes, general physiological deterioration, or mortality carried out comparisons with NEWS [42,45], MEWS [11,26,35,36], and the Thrombolysis in Myocardial Infarction score [36]. The 3 studies exploring sepsis-related outcomes additionally included the SOFA, qSOFA, and SIRS criteria and the simplified acute physiology (II) score [23,34,37]. A few studies also drew comparisons with other customized scoring systems individual to their care setting or region such as the Korean Triage and Acuity Score [11], Singapore Emergency Department Sepsis model [23], and postanesthesia care unit alarm system [46].

In all 9 studies, the ML models performed better than the aggregate-weighted EWS systems for all clinical outcomes except for cardiac arrest in the study by Badriyah et al [45]. For example, in the study by Jang et al [35], a long short-term memory neural network achieved an AUROC of 0.933, an improvement over MEWS, which achieved an AUROC of 0.886 using the same data. Similarly, in the study by Kwon et al [26], recurrent neural networks achieved an AUROC of 0.85 compared to 0.603 for MEWS and 0.785 for the Korean Triage and Acuity Score. Some studies reported much more modest improvements, such as the study by Chiu et al [42] that achieved

an AUROC of 0.779 using logistic regression, compared to 0.754 using MEWS for the same 24-hour prediction window. A full side-by-side comparison of ML vs aggregate-weighted EWS is presented in Multimedia Appendix 3.

Discussion

Based on this scoping review, ML-based EWS models show considerable promise, but there exist several important avenues for future research if these models are to be effectively implemented in clinical practice.

Prediction Window

A model's prediction window refers to how far in advance a model is predicting an adverse event. Most studies included in our review used a prediction window between 30 minutes [26] and 72 hours [36] before the clinical deterioration took place. The length of a model's prediction window is important because a prediction window that is too short will not yield any real clinical benefit (it would not give a clinical team sufficient time to intervene), but a number of studies [29,34,37,42] showed a decrease in model performance when the prediction window was longer (eg, AUROC drops from 0.88 at the time of onset to 0.74 at 4 hours before the event). Future research seeking to maximize the clinical benefit of ML EWS should strive to achieve an optimum balance between a clinically relevant prediction window and clinically acceptable model performance, rather than simply maximizing a model performance metric, such as AUROC.

Clinically Actionable Explanations

The studies included in this review focused on ML model development and did not explore how the output of these models would be communicated to clinicians. Since many ML models are "black boxes" [46,47], it may not be immediately clear to clinicians what the likely reason for an alert might be until the patient is assessed, which can cause further delays in time-sensitive scenarios. However, in the broader ML field, there has been significant recent progress in explainable ML techniques, and it has been pointed out that these approaches may be preferred by the medical community and regulators [48,49]. Several explanation methods take specific, previously black-box methods, such as convolutional neural networks [50], and allow for post-hoc explanation of their decision-making process. Other explainability algorithms are model-agnostic, meaning they can be applied to any type of model, regardless of its mathematical basis [51]. In the study by Lauritsen et al [52], an explainable EWS was developed based on a temporal convolutional network, using a separate module for explanations. These methodologies are promising, but their application to health care, including to EWS, has been limited. Objective evaluation of the utility of explanation methods is a difficult, ongoing problem, but is an important direction for future research in the area of ML-based EWS if they are to be effectively deployed in clinical practice [53].

Expanded Study Settings

Nearly all the studies included in this review were conducted in inpatient settings. While EWS are highly valuable in an inpatient context, there is also considerable need in the ambulatory setting, particularly postdischarge. For example, the VISION study [54] found that 1.8% of all patients die within 30 days postsurgery and 29.4% of all deaths occurred after patients were discharged from hospital. Patients often receive postoperative monitoring only 3-4 weeks [54] after discharge during a follow-up visit with their surgeon. During this period, it has been shown that many patients suffer from prolonged unidentified hypoxemia [55] and hypotension [56], which are precursors to serious postoperative complications. While EWS research has historically focused on inpatient settings due to the availability of continuous vital signs data, the increasing availability of remote patient monitoring and wearable technologies offer the opportunity to direct future EWS research to the ambulatory setting to address a significant clinical need.

Retrospective Versus Prospective Evaluation

All but one study [21] included in this review were retrospective in nature, leaving open the possibility that algorithm performance in a clinical environment may be lower than the performance achieved in a controlled retrospective setting [34]. It is also unclear how often these EWS were able to identify clinical deterioration that had not already been detected by a care team. Further, alerts for clinical deterioration may be easily disregarded by clinicians due to alert fatigue, even when the risk of deterioration has been correctly identified [43]. In the single case where an ML-based EWS was studied prospectively, Olsen et al [21] found that the random forest classifier decreased false alarm rates by 85% and the rate of missed alerts by 73% when compared to the existing aggregate-weighted alarm system. While the predictions were independently scored for severity by 2 clinician experts, the interpretation of the clinical impact of these alerts was not explored any further, leaving the question of clinical benefit unanswered. Future research into ML-based EWS should begin to include prospective evaluation, both of model accuracy (to understand how model performance is affected when faced with real-world data) and of clinical outcomes (to understand whether alerts in fact produce clinical benefits).

Standardizations of Performance Metrics

A key observation from this review is the lack of an agreed-upon standard among the research community for reporting performance measures across studies. This makes meaningful comparison between the outcomes of these studies difficult, and where there is overlap, it is not clear that the most clinically relevant metrics have been chosen. The majority of the studies in this review report the AUROC as the main performance metric, reflecting a common practice in the ML literature. However, AUROC may not be adequate for evaluating the performance of the EWS in a clinical setting [57].

As Romero-Brufau et al [58] discussed in their article, AUROC does not incorporate information about the prevalence of physiological deterioration, which can be lower than 0.02 daily in a general inpatient setting. This can make AUROC a misleading metric, leading to overestimation of clinical benefit and underestimation of clinical workload and resources. [58] When the prevalence is low (<0.1), even a model with high sensitivity and specificity may not yield a high posttest probability for a positive prediction [15]. Therefore, reporting

metrics that incorporate the prevalence would be more appropriate.

The performance of an EWS depends on the tradeoff between 2 goals: early detection of outcomes versus issuance of fewer false-positive alerts to prevent alarm fatigue [43]. Sensitivity can be a good metric to evaluate the first goal as it would provide the percentage of true-positive predictions within a certain time period. To evaluate the clinical burden of false-positive alerts, the positive predictive value, which incorporates prevalence, can be used as it gives a percentage of useful alerts that lead to a clinical outcome. The number needed to evaluate can be a useful measure of clinical utility and cost-efficiency of each alert as it provides the number of patients that need to be evaluated further to detect one outcome. Using these metrics to evaluate tradeoffs between outcome detection and workload would be essential for determining the clinical utility of the EWS [58]. Additionally, the F1 score can also be a useful metric as it provides a measure of the model's overall accuracy through the calculation of the harmonic mean of the precision and recall (sensitivity). Balancing the use of these 2 metrics could yield a more realistic measure of the model's performance [58].

Comparison to "Gold Standard" EWS

On a related note, only 9 of the studies included in our review made comparisons between their ML-based models and a "gold standard" aggregate-weighted EWS, such as MEWS or NEWS. Future research in the area should report a commonly used aggregate-weighted EWS as a baseline model, which would aid in making effective comparisons between them. NEWS may be particularly well suited to this area of research as its input variables can all be measured automatically and continuously via devices.

Strengths of the Review

The search strategy was comprehensive while not being too focused on specific clinical outcomes, sampling frequencies, or filtering for time. This allowed for the identification of as many studies as possible that examined the use of ML models and vital signs to predict the risk of patient deterioration. No additional studies were identified through citation tracking after the original search, indicating our search strategy was comprehensive. Unlike previous reviews, inclusion criteria for the review supported the examination of findings from studies conducted across a variety of clinical settings including specialty units or wards and ambulatory care. This helped in characterizing the use of ML-based prediction models in different patient-care environments with varying clinical endpoints. Wherever the original studies provided the data, comparisons were drawn between the performance of the ML models and that of aggregate-weighted EWS. This gives an indication of the differences in accuracy of the models in predicting clinical deterioration.

Limitations

The findings within this review are subject to some limitations. First, the literature search, assessment of eligibility of full-text articles, inclusion in the review, and extraction of study data were carried out by only 1 author. Second, only the findings from published studies were included in this scoping review, which may affect the results due to publication bias. While studies from a variety of settings were included, the generalizability of our findings may be limited due to the heterogeneity of patient populations, clinical practices, and study methodologies. Sampling procedures and frequencies varied across studies from single to multiple observations of patient vital signs, and clinical outcome definitions were based on different criteria or aggregate-weighted EWS. Finally, due to this variation in ML methods, prediction windows, and outcome reporting, a meta-analysis was not feasible.

Conclusion

Our findings suggest that ML-based EWS models incorporating easily accessible vital sign measurements are effective in predicting physiological deterioration in patients. Improved prediction performance was also observed with these models when compared to traditional aggregate-based risk stratification tools. The clinical impact of these ML-based EWS could be significant for clinical staff and patients due to decreased false alerts and increased early detection of warning signs for timely intervention, though further development of these models is needed and the necessary prospective research to establish actual clinical utility does not yet exist.

Authors' Contributions

SM contributed to conceptualization, data collection, data analysis, and manuscript writing. JP contributed to conceptualization, manuscript writing, and manuscript review. WN and SD contributed equally to manuscript writing and review. PD contributed to manuscript writing and review. MM and NB contributed to manuscript review.

Conflicts of Interest

PJD is a member of a research group with a policy of not accepting honorariums or other payments from industry for their own personal financial gain. They do accept honorariums/payments from industry to support research endeavours and costs to participate in meetings.

Based on study questions PJD has originated and grants he has written, he has received grants from Abbott Diagnostics, AstraZeneca, Bayer, Boehringer Ingelheim, Bristol-Myers-Squibb, Coviden, Octapharma, Philips Healthcare, Roche Diagnostics, Siemens, and Stryker.

PJD has participated in advisory board meetings for GlaxoSmithKline, Boehringer Ingelheim, Bayer, and Quidel Canada. He also attended an expert panel meeting with AstraZeneca and Boehringer Ingelheim.

The other authors declare no conflicts of interest.

Multimedia Appendix 1
Search terms.
[DOCX File , 12 KB - jmir_v23i2e25187_app1.docx]

Multimedia Appendix 2
Description of ML methods and relevant terms.
[DOCX File , 15 KB - jmir_v23i2e25187_app2.docx]

Multimedia Appendix 3
Comparison between performance of ML based EWS and aggregate EWS.
[DOCX File , 20 KB - jmir_v23i2e25187_app3.docx]

References

1. Barfod C, Lauritzen MMP, Danker JK, Sölétormos G, Forberg JL, Berlac PA, et al. Abnormal vital signs are strong predictors for intensive care unit admission and in-hospital mortality in adults triaged in the emergency department - a prospective cohort study. Scand J Trauma Resusc Emerg Med 2012 Apr 10;20:28 [FREE Full text] [doi: 10.1186/1757-7241-20-28] [Medline: 22490208]
2. Hillman KM, Bristow PJ, Chey T, Daffurn K, Jacques T, Norman SL, et al. Antecedents to hospital deaths. Intern Med J 2001 Aug;31(6):343-348. [doi: 10.1046/j.1445-5994.2001.00077.x] [Medline: 11529588]
3. McGaughey J, Alderdice F, Fowler R, Kapila A, Mayhew A, Moutray M. Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards. Cochrane Database Syst Rev 2007 Jul 18(3):CD005529. [doi: 10.1002/14651858.CD005529.pub2] [Medline: 17636805]
4. Smith GB, Prytherch DR, Schmidt PE, Featherstone PI, Higgins B. A review, and performance evaluation, of single-parameter "track and trigger" systems. Resuscitation 2008 Oct;79(1):11-21. [doi: 10.1016/j.resuscitation.2008.05.004] [Medline: 18620794]
5. Gardner-Thorpe J, Love N, Wrightson J, Walsh S, Keeling N. The value of Modified Early Warning Score (MEWS) in surgical in-patients: a prospective observational study. Ann R Coll Surg Engl 2006 Oct;88(6):571-575 [FREE Full text] [doi: 10.1308/003588406X130615] [Medline: 17059720]
6. Gao H, McDonnell A, Harrison DA, Moore T, Adam S, Daly K, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. Intensive Care Med 2007 Apr;33(4):667-679. [doi: 10.1007/s00134-007-0532-3] [Medline: 17318499]
7. Subbe CP, Slater A, Menon D, Gemmell L. Validation of physiological scoring systems in the accident and emergency department. Emerg Med J 2006 Nov;23(11):841-845 [FREE Full text] [doi: 10.1136/emj.2006.035816] [Medline: 17057134]
8. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. Resuscitation 2013 Apr;84(4):465-470. [doi: 10.1016/j.resuscitation.2012.12.016] [Medline: 23295778]
9. Tam B, Xu M, Kwong M, Wardell C, Kwong A, Fox-Robichaud A. The Admission Hamilton Early Warning Score (HEWS) Predicts the Risk of Critical Event during Hospitalization. Can Journ Gen Int Med 2017 Feb 24;11(4):1. [doi: 10.22374/cjgim.v11i4.190]
10. Prytherch DR, Smith GB, Schmidt PE, Featherstone PI. ViEWS--Towards a national early warning score for detecting adult inpatient deterioration. Resuscitation 2010 Aug;81(8):932-937. [doi: 10.1016/j.resuscitation.2010.04.014] [Medline: 20637974]
11. Kwon J, Lee Y, Lee Y, Lee S, Park H, Park J. Validation of deep-learning-based triage and acuity score using a large national dataset. PLoS One 2018;13(10):e0205836 [FREE Full text] [doi: 10.1371/journal.pone.0205836] [Medline: 30321231]
12. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood) 2014 Jul;33(7):1123-1131. [doi: 10.1377/hlthaff.2014.0041] [Medline: 25006137]
13. Churpek MM, Yuen TC, Park SY, Gibbons R, Edelson DP. Using electronic health record data to develop and validate a prediction model for adverse outcomes in the wards*. Crit Care Med 2014 Apr;42(4):841-848 [FREE Full text] [doi: 10.1097/CCM.0000000000000038] [Medline: 24247472]
14. Linnen DT, Escobar GJ, Hu X, Scruth E, Liu V, Stephens C. Statistical Modeling and Aggregate-Weighted Scoring Systems in Prediction of Mortality and ICU Transfer: A Systematic Review. J Hosp Med 2019 Mar;14(3):161-169 [FREE Full text] [doi: 10.12788/jhm.3151] [Medline: 30811322]
15. Brekke IJ, Puntervoll LH, Pedersen PB, Kellett J, Brabrand M. The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review. PLoS One 2019;14(1):e0210875 [FREE Full text] [doi: 10.1371/journal.pone.0210875] [Medline: 30645637]

16. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
17. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993 Apr;39(4):561-577. [Medline: 8472349]
18. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
19. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration. 2011. URL: www.handbook.cochrane.org [accessed 2021-01-09]
20. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg 2015 Feb;102(3):148-158. [doi: 10.1002/bjs.9736] [Medline: 25627261]
21. Olsen RM, Aasvang EK, Meyhoff CS, Dissing Sorensen HB. Towards an automated multimodal clinical decision support system at the post anesthesia care unit. Comput Biol Med 2018 Oct 01;101:15-21. [doi: 10.1016/j.compbiomed.2018.07.018] [Medline: 30092398]
22. Larburu N, Artetxe A, Escolar V, Lozano A, Kerexeta J. Artificial Intelligence to Prevent Mobile Heart Failure Patients Decompensation in Real Time: Monitoring-Based Predictive Model. Mobile Information Systems 2018 Nov 05;2018:1-11. [doi: 10.1155/2018/1546210]
23. Chiew CJ, Liu N, Tagami T, Wong TH, Koh ZX, Ong MEH. Heart rate variability based machine learning models for risk prediction of suspected sepsis patients in the emergency department. Medicine (Baltimore) 2019 Feb;98(6):e14197 [FREE Full text] [doi: 10.1097/MD.0000000000014197] [Medline: 30732136]
24. Churpek MM, Adhikari R, Edelson DP. The value of vital sign trends for detecting clinical deterioration on the wards. Resuscitation 2016 May;102:1-5 [FREE Full text] [doi: 10.1016/j.resuscitation.2016.02.005] [Medline: 26898412]
25. Clifton L, Clifton DA, Pimentel MAF, Watkinson PJ, Tarassenko L. Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors. IEEE J Biomed Health Inform 2014 May;18(3):722-730. [doi: 10.1109/JBHI.2013.2293059] [Medline: 24808218]
26. Kwon J, Lee Y, Lee Y, Lee S, Park J. An Algorithm Based on Deep Learning for Predicting In-Hospital Cardiac Arrest. J Am Heart Assoc 2018 Jun 26;7(13):1 [FREE Full text] [doi: 10.1161/JAHA.118.008678] [Medline: 29945914]
27. Forkan ARM, Khalil I. PEACE-Home: Probabilistic estimation of abnormal clinical events using vital sign correlations for reliable home-based monitoring. Pervasive and Mobile Computing 2017 Jul;38:296-311. [doi: 10.1016/j.pmcj.2016.12.009]
28. Forkan ARM, Khalil I. A clinical decision-making mechanism for context-aware and patient-specific remote monitoring systems using the correlations of multiple vital signs. Comput Methods Programs Biomed 2017 Feb;139:1-16. [doi: 10.1016/j.cmpb.2016.10.018] [Medline: 28187881]
29. Forkan ARM, Khalil I, Atiquzzaman M. ViSiBiD: A learning model for early discovery and real-time prediction of severe clinical events using vital signs as big data. Computer Networks 2017 Feb;113:244-257. [doi: 10.1016/j.comnet.2016.12.019]
30. Lee J, Scott D, Villarroel M, Clifford G, Saeed M, Mark R. Open-access MIMIC-II database for intensive care research. Conf Proc IEEE Eng Med Biol Soc 2011;2011:8315-8318. [doi: 10.1109/iembs.2011.6092050] [Medline: 22256274]
31. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
32. Tarassenko L, Hann A, Young D. Integrated monitoring and analysis for early warning of patient deterioration. Br J Anaesth 2006 Jul;97(1):64-68 [FREE Full text] [doi: 10.1093/bja/ael113] [Medline: 16707529]
33. van Wyk F, Khojandi A, Kamaleswaran R, Akbilgic O, Nemati S, Davis RL. How much data should we collect? A case study in sepsis detection using deep learning. 2017 Presented at: IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT); November 6-8, 2017; Bethesda, MD. [doi: 10.1109/hic.2017.8227596]
34. Mao Q, Jay M, Hoffman JL, Calvert J, Barton C, Shimabukuro D, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. BMJ Open 2018 Jan 26;8(1):e017833 [FREE Full text] [doi: 10.1136/bmjopen-2017-017833] [Medline: 29374661]
35. Jang D, Kim J, Jo YH, Lee JH, Hwang JE, Park SM, et al. Developing neural network models for early detection of cardiac arrest in emergency department. Am J Emerg Med 2020 Jan;38(1):43-49. [doi: 10.1016/j.ajem.2019.04.006] [Medline: 30982559]
36. Liu N, Koh ZX, Goh J, Lin Z, Haaland B, Ting BP, et al. Prediction of adverse cardiac events in emergency department patients with chest pain using machine learning for variable selection. BMC Med Inform Decis Mak 2014 Aug 23;14:75 [FREE Full text] [doi: 10.1186/1472-6947-14-75] [Medline: 25150702]
37. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. JMIR Med Inform 2016 Sep 30;4(3):e28 [FREE Full text] [doi: 10.2196/medinform.5909] [Medline: 27694098]
38. Ho J, Park Y. Learning from different perspectives: Robust cardiac arrest prediction via temporal transfer learning. Annu Int Conf IEEE Eng Med Biol Soc 2017 Jul;2017:1672-1675. [doi: 10.1109/EMBC.2017.8037162] [Medline: 29060206]

39. Li X, Wang Y. Adaptive online monitoring for ICU patients by combining just-in-time learning and principal component analysis. J Clin Monit Comput 2016 Dec;30(6):807-820. [doi: 10.1007/s10877-015-9778-4] [Medline: 26392184]
40. Shashikumar SP, Stanley MD, Sadiq I, Li Q, Holder A, Clifford GD, et al. Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. J Electrocardiol 2017;50(6):739-743 [FREE Full text] [doi: 10.1016/j.jelectrocard.2017.08.013] [Medline: 28916175]
41. Yoon JH, Mu L, Chen L, Dubrawski A, Hravnak M, Pinsky MR, et al. Predicting tachycardia as a surrogate for instability in the intensive care unit. J Clin Monit Comput 2019 Dec;33(6):973-985 [FREE Full text] [doi: 10.1007/s10877-019-00277-0] [Medline: 30767136]
42. Chiu Y, Villar SS, Brand JW, Patteril MV, Morrice DJ, Clayton J, et al. Logistic early warning scores to predict death, cardiac arrest or unplanned intensive care unit re-admission after cardiac surgery. Anaesthesia 2020 Feb;75(2):162-170 [FREE Full text] [doi: 10.1111/anae.14755] [Medline: 31270799]
43. Guillame-Bert M, Dubrawski A, Wang D, Hravnak M, Clermont G, Pinsky MR. Learning temporal rules to forecast instability in continuously monitored patients. J Am Med Inform Assoc 2017 Jan;24(1):47-53 [FREE Full text] [doi: 10.1093/jamia/ocw048] [Medline: 27274020]
44. Chen L, Ogundele O, Clermont G, Hravnak M, Pinsky MR, Dubrawski AW. Dynamic and Personalized Risk Forecast in Step-Down Units. Implications for Monitoring Paradigms. Ann Am Thorac Soc 2017 Mar;14(3):384-391 [FREE Full text] [doi: 10.1513/AnnalsATS.201611-905OC] [Medline: 28033032]
45. Badriyah T, Briggs JS, Meredith P, Jarvis SW, Schmidt PE, Featherstone PI, et al. Decision-tree early warning score (DTEWS) validates the design of the National Early Warning Score (NEWS). Resuscitation 2014 Mar;85(3):418-423. [doi: 10.1016/j.resuscitation.2013.12.011] [Medline: 24361673]
46. Cabitza F, Rasoini R, Gensini GF. Unintended Consequences of Machine Learning in Medicine. JAMA 2017 Aug 08;318(6):517-518. [doi: 10.1001/jama.2017.7797] [Medline: 28727867]
47. Stead WW. Clinical Implications and Challenges of Artificial Intelligence and Deep Learning. JAMA 2018 Sep 18;320(11):1107-1108. [doi: 10.1001/jama.2018.11029] [Medline: 30178025]
48. Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. Proceedings of the 4th Machine Learning for Healthcare Conference 2019;106:359-380.
49. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019 Sep;25(9):1337-1340. [doi: 10.1038/s41591-019-0548-6] [Medline: 31427808]
50. Wang ZJ, Turko R, Shaikh O, Park H, Das N, Hohman F, et al. CNN EXPLAINER: Learning Convolutional Neural Networks with Interactive Visualization. IEEE Trans. Visual. Comput. Graphics 2020:1. [doi: 10.1109/tvcg.2020.3030418]
51. Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Cornell University. 2016. URL: https://arxiv.org/abs/1602.04938 [accessed 2021-01-09]
52. Lauritsen SM, Kristensen M, Olsen MV, Larsen MS, Lauritsen KM, Jørgensen MJ, et al. Explainable artificial intelligence model to predict acute critical illness from electronic health records. Nat Commun 2020 Jul 31;11(1):3852 [FREE Full text] [doi: 10.1038/s41467-020-17431-x] [Medline: 32737308]
53. Sokol K, Flach P. Explainability Fact Sheets: A Framework for Systematic Assessment of Explainable Approaches. Cornell University. 2019. URL: https://arxiv.org/abs/1912.05100 [accessed 2021-01-09]
54. Vascular Events in Noncardiac Surgery Patients Cohort Evaluation (VISION) Study Investigators, Spence J, LeManach Y, Chan MT, Wang CY, Sigamani A, et al. Association between complications and death within 30 days after noncardiac surgery. CMAJ 2019 Jul 29;191(30):E830-E837 [FREE Full text] [doi: 10.1503/cmaj.190221] [Medline: 31358597]
55. Sun Z, Sessler DI, Dalton JE, Devereaux PJ, Shahinyan A, Naylor AJ, et al. Postoperative Hypoxemia Is Common and Persistent: A Prospective Blinded Observational Study. Anesth Analg 2015 Sep;121(3):709-715 [FREE Full text] [doi: 10.1213/ANE.0000000000000836] [Medline: 26287299]
56. Turan A, Chang C, Cohen B, Saasouh W, Essber H, Yang D, et al. Incidence, Severity, and Detection of Blood Pressure Perturbations after Abdominal Surgery: A Prospective Blinded Observational Study. Anesthesiology 2019 Apr;130(4):550-559 [FREE Full text] [doi: 10.1097/ALN.0000000000002626] [Medline: 30875354]
57. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 2007 Feb 20;115(7):928-935 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.106.672402] [Medline: 17309939]
58. Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. Crit Care 2015 Aug 13;19:285 [FREE Full text] [doi: 10.1186/s13054-015-0999-1] [Medline: 26268570]

Abbreviations

AUROC: area under the receiver operating characteristic
AVPU: alert, verbal, pain, unresponsive
BP: blood pressure
ED: emergency department
EWS: early warning system

HR: heart rate
ICU: intensive care unit
MEWS: Modified Early Warning Score
MIMIC: Medical Information Mart for Intensive Care
ML: machine learning
NEWS: National Early Warning Score
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
qSOFA: quick Sequential Organ Failure Assessment
RR: respiratory rate
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

Edited by G Eysenbach; submitted 21.10.20; peer-reviewed by N Liu, J Kellett; comments to author 07.11.20; revised version received 19.12.20; accepted 20.12.20; published 04.02.21.

Please cite as:
Muralitharan S, Nelson W, Di S, McGillion M, Devereaux PJ, Barr NG, Petch J
Machine Learning–Based Early Warning Systems for Clinical Deterioration: Systematic Scoping Review
J Med Internet Res 2021;23(2):e25187
URL: https://www.jmir.org/2021/2/e25187
doi:10.2196/25187
PMID:33538696

Review

Digital Health Interventions for Cardiac Rehabilitation: Systematic Literature Review

Shannon Wongvibulsin[1], PhD; Evagelia E Habeos[2], MD; Pauline P Huynh[1], BA; Helen Xun[1], BS; Rongzi Shan[3,4], BS; Kori A Porosnicu Rodriguez[1], BA; Jane Wang[1,4], MD; Yousuf K Gandapur[5], MD; Ngozi Osuji[3], MD, MPH; Lochan M Shah[1], MD; Erin M Spaulding[6], PhD, RN, BSN; George Hung[1], MD; Kellen Knowles[7], MD; William E Yang[1], MD; Francoise A Marvel[3], MD; Eleanor Levin[8], MD, FACC, FAHA; David J Maron[9], MD, FACC, FAHA; Neil F Gordon[10,11], MD, PhD, MPH; Seth S Martin[3], MD, MHS, FACC, FAHA, FASPC

[1]Johns Hopkins University School of Medicine, Baltimore, MD, United States

[2]University of Patras School of Medicine, Patras, Greece

[3]Ciccarone Center for the Prevention of Cardiovascular Disease, Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, United States

[4]UCLA David Geffen School of Medicine, Los Angeles, CA, United States

[5]Medstar Franklin Square Hospital, Baltimore, MD, United States

[6]Johns Hopkins School of Nursing, Baltimore, MD, United States

[7]Department of Medicine, Johns Hopkins Bayview Medical Center, Baltimore, MD, United States

[8]Department of Medicine, Division of Cardiology, Stanford University School of Medicine, Stanford, CA, United States

[9]Stanford Prevention Research Center, Stanford University School of Medicine, Stanford, CA, United States

[10]INTERVENT International, Savannah, GA, United States

[11]Centre for Exercise Science and Sports Medicine, School of Therapeutic Sciences, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

Corresponding Author:
Seth S Martin, MD, MHS, FACC, FAHA, FASPC
Ciccarone Center for the Prevention of Cardiovascular Disease, Division of Cardiology
Department of Medicine
Johns Hopkins University School of Medicine
600 N Wolfe Street
Carnegie 591
Baltimore, MD, 21287
United States
Phone: 1 410 502 0469
Email: smart100@jhmi.edu

Abstract

Background: Cardiovascular disease (CVD) is the leading cause of death worldwide. Despite strong evidence supporting the benefits of cardiac rehabilitation (CR), over 80% of eligible patients do not participate in CR. Digital health technologies (ie, the delivery of care using the internet, wearable devices, and mobile apps) have the potential to address the challenges associated with traditional facility-based CR programs, but little is known about the comprehensiveness of these interventions to serve as digital approaches to CR. Overall, there is a lack of a systematic evaluation of the current literature on digital interventions for CR.

Objective: The objective of this systematic literature review is to provide an in-depth analysis of the potential of digital health technologies to address the challenges associated with traditional CR. Through this review, we aim to summarize the current literature on digital interventions for CR, identify the key components of CR that have been successfully addressed through digital interventions, and describe the gaps in research that need to be addressed for sustainable and scalable digital CR interventions.

Methods: Our strategy for identifying the primary literature pertaining to CR with digital solutions (defined as technology employed to deliver remote care beyond the use of the telephone) included a consultation with an expert in the field of digital CR and searches of the PubMed (MEDLINE), Embase, CINAHL, and Cochrane databases for original studies published from January 1990 to October 2018.

Results: Our search returned 31 eligible studies, of which 22 were randomized controlled trials. The reviewed CR interventions primarily targeted physical activity counseling (31/31, 100%), baseline assessment (30/31, 97%), and exercise training (27/31, 87%). The most commonly used modalities were smartphones or mobile devices (20/31, 65%), web-based portals (18/31, 58%), and email-SMS (11/31, 35%). Approximately one-third of the studies addressed the CR core components of nutrition counseling, psychological management, and weight management. In contrast, less than a third of the studies addressed other CR core components, including the management of lipids, diabetes, smoking cessation, and blood pressure.

Conclusions: Digital technologies have the potential to increase access and participation in CR by mitigating the challenges associated with traditional, facility-based CR. However, previously evaluated interventions primarily focused on physical activity counseling and exercise training. Thus, further research is required with more comprehensive CR interventions and long-term follow-up to understand the clinical impact of digital interventions.

(J Med Internet Res 2021;23(2):e18773) doi:10.2196/18773

KEYWORDS

cardiac rehabilitation; telemedicine; digital technologies; mHealth; mobile phone

Introduction

Cardiac Rehabilitation

Cardiovascular disease (CVD) is the leading cause of death worldwide, with approximately 80% of CVD resulting from modifiable risk factors such as physical inactivity, poor dietary habits, elevated low-density lipoprotein-cholesterol and plasma glucose levels, and smoking [1]. Following a cardiac event, cardiac rehabilitation (CR) is an effective modality that enhances recovery, reduces cardiovascular mortality and risk for hospital admissions, and improves the health-related quality of life (QoL) [2]. CR is a multi-faceted, medically supervised program that addresses established core components of guideline-directed therapy, including baseline patient assessments, nutritional counseling, risk factor modification (including management of lipids, blood pressure, weight, diabetes mellitus, and smoking), psychosocial interventions, and physical activity counseling and exercise training [3]. Although there is strong evidence supporting the benefits of CR, less than 20% of patients who are eligible participate in CR [4]. Challenges related to the low utilization of CR include the lack of referral or facilitation of enrollment, limited health insurance coverage, time and costs associated with participation and travel, and lack of access to a CR facility because of scheduling, transportation, or distance [5].

Digital Technology for CR

The technology for CR is advancing rapidly and has the potential to address the challenges of traditional facility-based CR programs by delivering care to patients in the convenience of their own homes with real-time, personalized support. As noted in the literature, the terminology describing this technology has not been standardized and includes *telemedicine*, *telehealth*, and *eHealth* [6,7]. In this review, we use the term *digital health interventions* to encompass technology that enables the delivery of care through means such as the use of the internet, wearable devices, and mobile apps [8,9]. Although there have been encouraging results from the use of digital health interventions for CR (eg, remote electrocardiographic monitoring and mobile or web portal tools), these developments have largely remained in the research settings and have not yet translated into widespread use in clinical practice [3]. Currently, there are gaps in understanding the comprehensiveness of digital CR programs and how successful they are in addressing the core components of CR. To help guide the development of digital CR interventions that have the potential to translate into clinical use, we have focused on the evaluation of technology used in digital interventions for CR and the comprehensiveness of these programs using the framework outlined in the scientific statement from the American Heart Association (AHA) and the American Association of Cardiovascular and Pulmonary Rehabilitation (AACVPR) for the core components of CR [3]. The specifics regarding the accreditation of CR programs are beyond the scope of this review. With the increasing need for technological advancements to revolutionize the delivery of CR care, this systematic literature review: (1) summarizes the current literature on digital interventions for CR, (2) identifies the key components of CR that have been successfully addressed through digital interventions, and (3) describes the gaps in research that need to be addressed for the sustainable implementation of digital CR interventions in clinical practice.

Methods

Overview

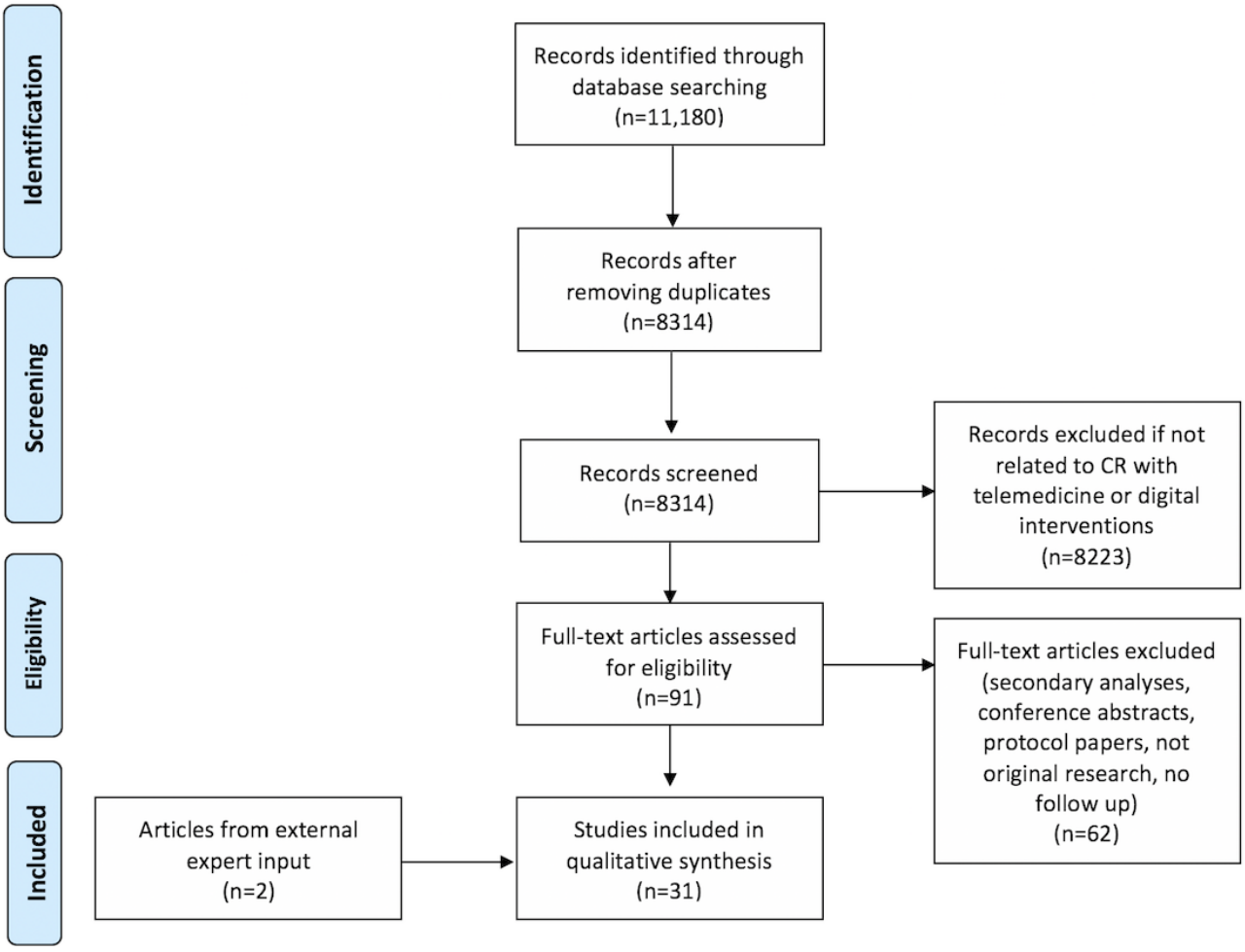
We designed a systematic, thematic review to answer key questions regarding the study designs to evaluate CR interventions, technology used, study size, and comprehensiveness of the investigated interventions. A full list of questions is provided in Textbox 1. Our search terms are detailed in Multimedia Appendix 1. We searched the PubMed (MEDLINE), Embase, CINAHL, and Cochrane databases for studies on digital CR published in English between January 1, 1990, and October 18, 2018. For this review, *digital* is defined as technology employed to deliver remote care beyond the use of telephone (eg, the delivery of care using the internet, wearable devices, and mobile apps). Telephonic-only studies, which have been addressed in the 2019 scientific statement on home-based CR from the AACVPR, AHA, and American College of Cardiology (ACC) [10], are not within the scope of this review. To determine eligibility for inclusion in this study, titles and abstracts were screened for relevance before a full-text review. The inclusion criteria for this review were as follows: (1) original research study using digital or telemedicine approaches for CR and (2) reported results for feasibility, usability, or

clinical outcomes. Studies were excluded if they (1) were not full-length publications (ie, abstracts), (2) were methods papers, (3) described only the technology without any inclusion of study participants, or (4) did not include any follow-up time to study outcomes (ie, cross-sectional studies). Given the evolving terminology surrounding digital health technology, we sought external expert inputs to include articles that were not found through our primary search strategy. Papers were included if they reported original research in digital CR. Full details are presented in Figure 1.

Textbox 1. Key questions to evaluate digital cardiac rehabilitation programs.

1. Which study designs were employed to evaluate the digital cardiac rehabilitation (CR) interventions?
2. Which technologies were used?
3. In which countries were these studies performed?
4. What were the study sample sizes?
5. What were the durations of the interventions and follow-up times?
6. What were the findings of these digital CR intervention studies?
7. How comprehensive were the digital CR interventions?

Figure 1. Flowchart for study identification, screening, eligibility, and inclusion. CR: cardiac rehabilitation.



Evaluation of CR Components and Study Quality

For each study, we recorded the components of CR that were delivered as described in the AHA and AACVPR consensus statement on the core components of CR and categorized the digital intervention listed in each study as either standalone or adjunctive to conventional CR. Studies were designated as standalone interventions when the program was delivered remotely with the exception of initial in-person session(s) for onboarding or baseline or outcome assessments, as long as the rest of the intervention was remote. The quality of the articles was independently assessed by 2 evaluators using the National Institutes of Health: National Heart, Lung, and Blood Institute (NHI: NHLBI) Study Quality Assessment Tools, which include the evaluation of 14 criteria for an overall quality assessment of good, fair, or poor (Multimedia Appendices 2 and 3) [11]. Discrepancies in ratings were resolved by discussions between the evaluators to reach a consensus on the ratings. We followed

the guidelines outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses [12].

Results

Study Characteristics

In total, 31 studies met the eligibility criteria and were included in this review (Table 1). The study characteristics are summarized in Table 2 and Multimedia Appendix 4. The median sample size was 98 (IQR 52.5-146), the median intervention duration was 3 months (IQR 1.6-4.4 months), and the median follow-up time was 6 months (IQR 3-6 months). The majority of these studies were conducted in Europe (12/31, 39%) and North America (8/31, 26%). A total of 22 studies (22/31, 71%) were randomized controlled trials. Of the 31 studies, 15 (15/31, 48%) were standalone digital CR interventions. The study quality was variable, with 23 studies (23/31, 74%) having a good quality, 7 (7/31, 23%) having a fair quality, and 1 (1/31, 3%) having a poor quality, according to the criteria established by the NIH: NHLBI Quality Assessment Tools (Multimedia Appendix 5).

As shown in Figure 2, the most commonly targeted CR core components were physical activity counseling (31/31, 100%), baseline assessment (30/31, 97%), and exercise training (27/31, 87%). Only about one-third of the studies addressed each of the other CR core components of nutrition counseling (11/31, 35%), psychological management (11/31, 35%), and weight management (10/31, 32%). In contrast, less than a third of the studies addressed other CR core components, with only a single study including lipid management (1/31, 3%), 2 studies including diabetes management (2/31, 6%), 7 studies including tobacco cessation (7/31, 23%), and 8 studies including blood pressure management (8/31, 26%). Smartphones/mobile devices and wearables were employed in 65% (20/31) of the studies, websites or web portals in 58% (18/31), and email-SMS communications in 35% (11/31) of the studies. The interventions were most commonly guided by physical therapists or exercise specialists (12/31, 39%), followed by CR/research team staff (11/31, 35%), and nurses (10/31, 32%). Four studies (4/31, 13%) described their interventions as fully automated or did not indicate requirement of any specific personnel [13-16]. The most commonly evaluated outcome was exercise capacity or step count (22/31, 71%). Other frequently assessed outcomes included program adherence (14/31, 45%) and QoL (14/31, 45%).

Table 1. Characteristics of the included studies.

Reference	Quality	Country	Design and population	Intervention	CR[a] components delivered[b]	Personnel and delivery setting[c]	Key outcomes
Ades, 2000 [17]	Fair	United States	Nonrandomized trial of patients with ACS[d] within past 3 months	3-month home-based, transtelephonically monitored CR, compared with conventional CR	• BA[e] • PAC[f] • ET[g]	• Nurse coordinator • IPS[h] • SAI[i]	• Exercise capacity • QoL[j]
Jenny, 2001 [13]	Fair	China	RCT[k] among cardiac patients enrolled in CR	30-min interactive computer-based health education program, compared to conventional health tutorial sessions	• BA • PAC	• Personnel required not specified • Program delivered through desktop or laptop computer	• Exercise self-efficacy • Exercise knowledge

Gordon, 2002 [18]	Good	United States	RCT among CAD[l] patients	12-week physician-supervised, nurse-case-managed cardiovascular risk reduction program and a community-based cardiovascular risk reduction program (including counseling via the telephone and internet) to patients with low-to-moderate-risk CAD as compared to contemporary phase II CR program	• BA
					• NC[m]
					• WM[n]
					• LM[o]
					• TC[p]
					• PAC
					• ET
					• PM[q]
						• Physician-supervised program: physician, nurse-case manager
						  • IPS
						• Community-based program: exercise physiologists, non-physician
						  • Health care professionals
						  • Physicians
						  • IPS
							• Maximal oxygen uptake
							• BP
							• Weight
							• Lipid profile
							• Medication use

Southard, 2003 [19]	Good	United States	RCT among CVD[r] patients	6-month internet-based program containing risk factor management support, education, and monitoring services to patients with CVD, as compared to usual care	• BA • NC • BPM • PAC • ET	• Case manager, dietician • IPS • SAI	• Satisfaction • Participation • Cost-effectiveness • Weight • BP • Lipid profile • Depression • Exercise capacity • Dietary habits
Barnason, 2009 [14]	Good	United States	RCT among CABG[s] patients	6-week symptom management telehealth intervention comprised of questionnaires, accelerometer, activity diary compared to standard of care	• BA • PAC • ET	None specified	Exercise capacity

Reference	Quality	Country	Design and population	Intervention	CR[a] components delivered[b]	Personnel and delivery setting[c]	Key outcomes
Scalvini, 2009 [20]	Good	Italy	Pilot study of patients with postop CABG or valve surgery	1-month home-based CR with remotely transmitted ECGs[t]	• BA • PAC • ET	• Nurse-tutor, physiotherapist • GS[u] • IPS • SAI	Step count
Piotrowicz, 2010 [21]	Fair	Poland	RCT among patients with HF[v]	8-week home-based telemonitored CR, compared with conventional CR	• BA • WM • BPM[w] • PAC • ET • PM	• Physician, physiotherapist • ECG technician, psychologist • IPS • SAI	• Intervention adherence • Exercise capacity
Reid, 2011 [22]	Good	Canada	RCT among CHD patients not participating in CR	6-month web-based tailored exercise intervention with email coaching, compared to standard of care	• BA • PAC • ET	• Exercise specialist • IPS • SAI	Exercise capacity
Clark, 2013 [23]	Fair	Australia	Pilot study of patients with post-MI[x] or angioplasty	7-week web-based CR intervention with educational materials, workbooks, and discussion forums, glucose and BP[y] monitoring, and pedometer	• BA • NC • WM • BPM • DM[z] • TC • PAC • PM	• General practitioner, nurse, allied health professional, case manager • IPS • SAI	Engagement
Brough, 2014 [24]	Fair	United Kingdom	Pilot study of patients with CHD[aa] referred for CR	8-week web-based CR comprised of web-based coaching and exercise e-diary	• BA • NC • WM • TC • PAC • ET • PM	• CR specialist • IPS • SAI	• Exercise capacity • Nutrition • Psychosocial well-being
Devi, 2014 [25]	Good	England	RCT among patients with CHD	6-week web-based CR comprised of exercise diary and web-based coaching, compared to standard of care	• BA • NC • WM • TC • PAC • ET • PM	• Researcher • IPS • SAI	• Step count • Exercise capacity • Weight • BP • Body fat percentage • QoL measures • Self-efficacy • Anxiety or depression • Dietary habits
Forman, 2014 [26]	Good	United States	Pilot study of patients enrolled in CR		• PAC • ET • PM	Nurse manager, exercise physiologist, nutritionist	Engagement

Reference	Quality	Country	Design and population	Intervention	CR[a] components delivered[b]	Personnel and delivery setting[c]	Key outcomes
				30-day task-based smartphone CR intervention comprised of medication and walking reminders, surveys, and educational tools, with web-based monitoring			
Kraal, 2014 [27]	Good	Netherlands	RCT among low- to moderate-risk CR patients	12-week home-based CR with telemonitored coaching interventions compared to standard of care	• BA • PAC • ET	• Physical therapist • IPS	• Exercise capacity • QoL
Piotrowicz, 2014 [28]	Poor	Poland	Nonrandomized trial of CVD patients referred for outpatient phase II CR	4-week home-based CR with remote ECG monitoring with mobile phone transmission	• BA • PAC • ET	• Nurse • IPS • SAI	• Intervention adherence • Satisfaction • Exercise capacity
Varnfield, 2014 [29]	Fair	Australia	RCT among post-MI patients referred to CR	6-week home-based CR using smartphone interventions (educational materials, exercise monitoring, weekly coaching), compared with conventional CR	• BA • NC • WM • BPM • TC • PAC • ET • PM	• Mentor (health coach) • IPS • SAI	• Intervention adherence • QoL • Exercise capacity • Weight
Whittaker, 2014 [30]	Fair	Australia	RCT among patients at post-MI	6-week home telehealth-based CR comprising mobile phone, Wellness Diary and web portal with tele-coaching, as compared to hospital-based CR	• BA • PAC • ET	• Health coach • IPS • SAI	• Health outcomes • Efficacy • Participation • Cost-effectiveness
Pfaeffli Dale, 2015 [16]	Good	New Zealand	Qualitative survey of patients with CHD	24-week mobile health program comprising text messaging and web-based coaching plus center-based CR, compared to center-based CR alone	• BA • NC • TC • PAC • PM	• Fully automated digital intervention • IPS	• Lifestyle modification • QoL • Intervention adherence
Frederix, 2015 [31]	Good	Belgium	RCT among CAD patients who completed phase II CR	18-week telemonitored exercise program, compared with standard of care	• BA • PAC • ET	• Rehabilitation center staff • IPS	• Exercise capacity • Weight • Lipid profile • Glycemic control • Rehospitalizations

Reference	Quality	Country	Design and population	Intervention	CR[a] components delivered[b]	Personnel and delivery setting[c]	Key outcomes
Lear, 2015 [32]	Good	Canada	RCT among patients with ACS or postrevascularization	4-month web-based CR comprising education, coaching, and physiologic data monitoring, compared with standard of care	• BA • NC • WM • BPM • DM • PAC • ET	• Program nurse, case manager, exercise specialist, dietician • GS • IPS • SAI	• Exercise capacity • Lipid profile • Dietary outcomes
Maddison, 2015 [15]	Good	New Zealand	RCT among patients with IHD[ab]	24-week smartphone-based intervention (website, educational videos, text messaging) plus standard of care, compared with standard of care alone	• BA • NC • PAC • ET • PM	Not specified	• Exercise capacity • QoL • Cost-effectiveness
Smolis-Bak, 2015 [33]	Good	Poland	Prospective randomized study among patients with HF and implanted CRT-D[ac]	8-week telemonitored home-based CR, compared to no training program after discharge	• BA • PAC • ET	• CR center staff, physiotherapist, doctor, nurse • IPS	• Exercise capacity • Echo evaluation • QoL
Frederix, 2016 [34]	Good	Belgium	Cost-effectiveness analysis of patients with CR	24-week web-based telerehabilitation program (web-based coaching, accelerometer) plus CR, compared to CR alone	• BA • NC • TC • PAC • ET	• Cardiac nurse, rehabilitation nurse • IPS	• Cost-effectiveness • Rehospitalizations
Skobel, 2016 [35]	Good	• Germany • Spain • United Kingdom	RCT among patients with CAD referred for CR	6-month smartphone-based exercise intervention (remote monitoring, physiologic data capture, and coaching), as compared to conventional CR	• BA • PAC • ET	• Sports physicians, exercise scientists • IPS • SAI	Exercise capacity
Thorup, 2016 [36]	Good	Denmark	RCT among hospitalized patients with ACS, HF, or coronary bypass surgery	3-month telerehabilitation trial with pedometer, compared among 3 rehabilitation settings	• BA • WM • BPM • PAC	• Personal nurse • GS • IPS	Step count
da Silva Vieira, 2017 [37]	Good	Portugal	RCT among patients who completed CR	6-month virtual reality CR intervention (Kinect) or booklet CR intervention, compared with standard of care	• BA • PAC • ET	• Researchers • IPS	• Body composition • Eating patterns • Lipid profile
	Good	Australia					

Reference	Quality	Country	Design and population	Intervention	CR[a] components delivered[b]	Personnel and delivery setting[c]	Key outcomes
Hwang, 2017 [38]			RCT among stable patients with chronic HF	12-week home-based CR with web-based video conferencing, compared with facility-based CR	• BA • NC • PAC • ET • PM	• Research staff, physiotherapists • GS • IPS • SAI	• Exercise capacity • QoL
Fang, 2018 [39]	Good	China	RCT among patients at post-PCI[ad]	6-week home-based CR with remote physiological monitoring and education, as compared with conventional CR	• BA • PAC • ET	• Medical team • IPS	• Exercise capacity • BP • QoL • Nicotine dependence
Harzand, 2018 [40]	Good	United States	Pilot study of veterans with CHD and eligible for CR	12-week home-based CR with smartphone app utilizing exercise reminders, educational materials, vitals monitoring, and remote coaching	• BA • WM • BPM • PAC • ET	• CR coach (cardiology physician assistant) • SAI	• Feasibility • BP • Acceptability • Exercise capacity
Maddison, 2018 [41]	Good	New Zealand	RCT among patients with CHD	12-week remotely monitored telerehabilitation with coaching, compared with conventional CR	• BA • PAC • ET	• CR exercise specialist • IPS	• Exercise capacity • QoL • Intervention adherence
Peng, 2018 [42]	Good	China	RCT among patients with HF	8-week home-based CR with remote coaching using physiologic data capture, web-based portal, and smartphone, compared with standard of care	• BA • WM • BPM • PAC • ET • PM	• Multidisciplinary CR team • IPS	• Exercise capacity • QoL • Echo evaluation
Rawstorn, 2018 [43]	Good	New Zealand	RCT among patients with CHD eligible for CR	12-week remotely monitored telerehabilitation with coaching, as compared with conventional CR	• BA • PAC • ET	Exercise specialist	• Usability • Satisfaction

[a]CR: cardiac rehabilitation.

[b]CR components were delivered through digital interventions except for baseline assessment that were conducted in person.

[c]Delivery setting: group sessions (GS), in-person session (IPS), standalone intervention (SAI).

[d]ACS: acute coronary syndrome.

[e]BA: baseline assessment.

[f]PAC: physical activity counseling.

[g]ET: exercise training.

[h]IPS: in-person session.

[i]SAI: standalone intervention.

[j]QoL: quality of life.

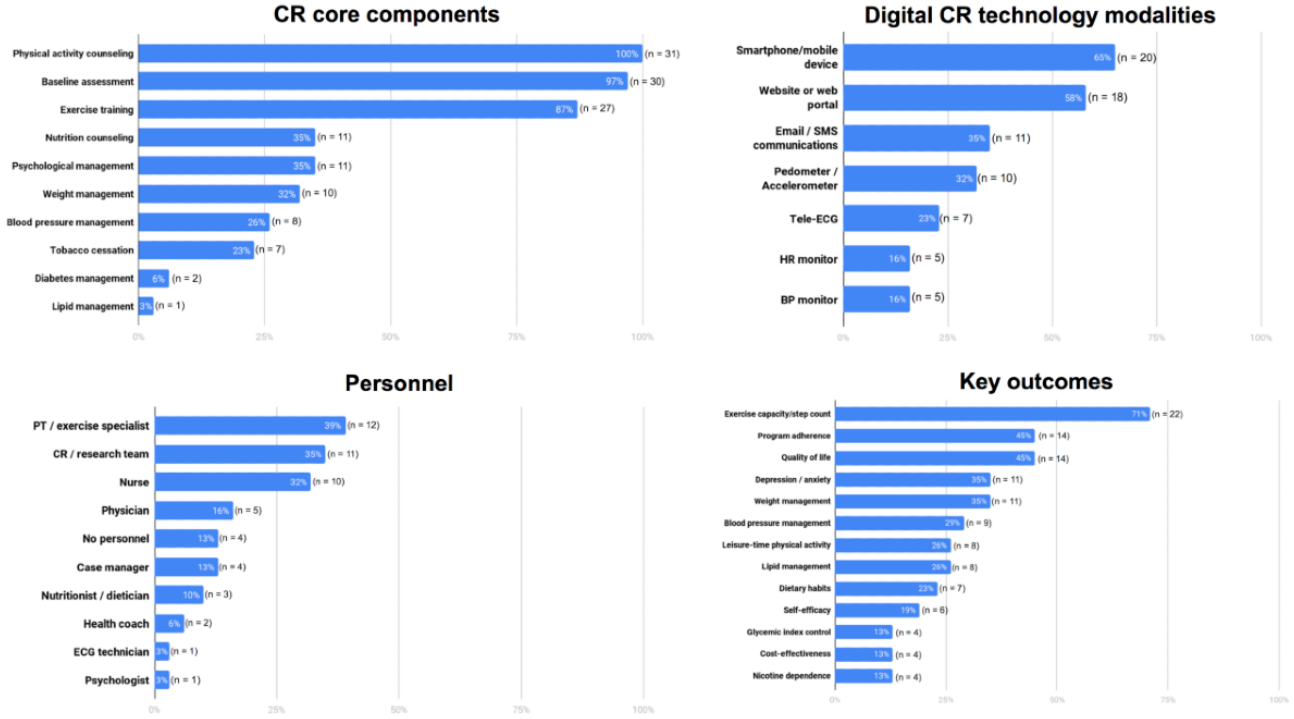
[k]RCT: randomized controlled trial.
[l]CAD: coronary artery disease.
[m]NC: nutrition counseling.
[n]WM: weight management.
[o]LM: lipid management.
[p]TC: tobacco cessation.
[q]PM: psychological management.
[r]CVD: cardiovascular disease.
[s]CABG: coronary artery bypass grafting.
[t]ECG: electrocardiogram.
[u]GS: group session.
[v]HF: heart failure.
[w]BPM: blood pressure management.
[x]MI: myocardial infarction.
[y]BP: blood pressure.
[z]DM: diabetes management.
[aa]CHD: coronary heart disease.
[ab]IHD: ischemic heart disease.
[ac]CRT-D: cardiac resynchronization therapy with defibrillator function.
[ad]PCI: percutaneous coronary intervention.

Table 2. Summary of the studies included in the analysis (n=31).

Characteristics	Value
Type of study, n (%)	
RCT[a]	22 (71)
Pilot study	5 (16)
Nonrandomized trial	2 (6)
Cost-effectiveness analysis from RCT	1 (3)
Qualitative study	1 (3)
Location of study, by continent, n (%)	
Europe	12 (39)
North America	8 (26)
Australia (including New Zealand)	8 (26)
Asia	3 (10)
Publication year, n (%)	
1990-2000	1 (3)
2001-2010	6 (19)
2011-2018	24 (77)
Median sample size (IQR)	98 (52.5-146)
Median follow-up time (months; IQR)	6 (3-6)
Median intervention duration (months; IQR)	3 (1.6-4.4)

[a]RCT: randomized controlled trial.

Figure 2. Percentage of studies with (1) the cardiac rehabilitation (CR) core components being addressed through digital interventions (apart from baseline assessment, which was conducted in person), (2) the technology modalities used in digital CR interventions, (3) the types of personnel employed in each CR program, and (4) the key outcomes evaluated. Cumulative percentages in some instances exceed 100% because some studies looked at multiple traits (n=31). BP: blood pressure; CR: cardiac rehabilitation; ECG: electrocardiogram; HR: heart rate; PT: physical therapist.



Key Outcomes

The key findings from the 31 studies are summarized in Textbox 2. Overall, program adherence was greater in patients using digital interventions than in those participating in conventional CR. Moreover, digital CR interventions were comparable to conventional CR (control) groups across multiple short-term outcomes (eg, functional capacity, physical activity, self-efficacy, program adherence, weight management, dietary habits, and QoL). However, digital CR interventions had mixed efficacy with regard to blood pressure control and mood. Of the 9 studies reporting blood pressure as an outcome of digital interventions versus conventional CR, 3 studies showed noninferiority [18,29,41], 4 found no significant impact on blood pressure [19,28,31,39], and 2 reported a better control of blood pressure in the control group as compared to the digital intervention group [25,35]. Similarly, although the majority of studies assessing mood as an outcome reported that digital interventions were noninferior to conventional CR in improving mood [22,25,29,30,33], 4 studies reported no significant improvement in anxiety or depression [19,24,25,42].

Textbox 2. Summary of findings by thematic outcomes.

Blood glucose control

- Only 1 of the 4 studies reported an improvement in glycemic control in the intervention group [31]. In total, 2 studies found no significant impact of the digital cardiac rehabilitation (CR) intervention on glycemic control [29,35]. One study found no significant difference in glycemic control between the digital intervention and usual care groups [41].

Blood pressure management

- In total, 3 of the 9 studies described CR interventions that significantly improved blood pressure management and were noninferior to the control groups [18,29,41]. A total of 4 studies did not find the digital CR interventions to significantly influence blood pressure management [19,28,31,39]. In all, 2 studies found that the control group had better blood pressure management as compared to the intervention group [25,35].

Depression or anxiety

- In total, 5 studies reported a positive effect on mood in the intervention group [22,25,29,30,33]. A total of 4 studies found that digital interventions had no significant impact on mood [19,24,25,42]. In total, 2 studies found no significant change in the psychological status between digital intervention and usual care groups [35,39]. One study found that both the intervention and usual care groups experienced an improvement in mood as compared with baseline, and there was no significant difference between the intervention and control groups [39]. One study found a negative effect on depression in the intervention group [16]. Of note, Devi et al [25] found that mood improved in the short term but was not significantly different from baseline at 6 months.

Dietary habits

- In total, 5 of the 7 studies found improvement in dietary habits [24,29,30,32,37], whereas 2 studies found that the intervention had no significant impact on the participants' dietary habits [19,25].

Exercise capacity

- In total, 22 studies looked at functional capacity as an outcome, and all of them reported that the intervention group was not inferior to the control group [15,17-22,24,25,27-33,35,38-42]. Of note, 4 studies found no significant difference in the functional capacity between the digital intervention and usual care groups [19,31,33,41]. Furthermore, Smolis-Bak et al [33] found functional capacity at the 12-month follow-up to be comparable to that at the baseline visit.

Lipid management

- In total, 2 of the 8 studies reported an increase in high-density lipoprotein in the intervention group [31,37], 2 of the 8 reported decreased low-density lipoprotein and total cholesterol [18,32], and 2 of the 8 studies reported a decrease in triglycerides [29,30].

Nicotine dependence

- Only 1 [39] of the 4 studies showed an improvement in smoking habits as measured by the Fagerstrom Test for Nicotine Dependence score [18,19,24,39].

Physical activity

- All 8 studies examining physical activity found that there was an improvement in the intervention group, comparable to or even greater than the increase in the physical activity in the control group [14,15,22,25,31,32,36,41].

Program adherence

- All 14 studies evaluating program adherence to CR reported that adherence to digital interventions was not inferior to traditional interventions [19-21,26-30,35,37,38,40,41,43]. Of note, 9 of the 14 studies found adherence to be greater in the digital intervention group [20,21,26,28-30,35,38,40].

Quality of life (QoL)

- In total, 10 studies reported an improvement in QoL in the intervention group [15,17,22,24,25,27,29,33,39,42]. A total of 6 studies found no significant difference in QoL measures between the digital intervention and usual care groups [17,21,27,35,38,41].

Self-efficacy

- Of the 6 studies evaluating self-efficacy, 5 showed improvement in self-efficacy following the digital interventions [13,15,25,32,41]. Pfaeffli Dale et al [16] found that the digital CR intervention had no significant impact on the participants' overall self-efficacy.

Weight management

- Digital CR interventions effectively addressed weight management in 8 of the 11 studies [18,19,25,28-30,37,41]. A total of 2 studies found no difference in weight or body mass index before and after the digital CR intervention [24,35]. In one study [17], the home group had increased weight, whereas the on-site control group had a slightly decreased weight.

We found a paucity of studies specifying intervention components that targeted lipid management, glycemic index control, and smoking cessation. These components were often reported as secondary outcome measures, if at all. Furthermore, the research on the long-term efficacy of digital CR interventions was sparse with heterogeneous findings. Although Devi et al [25] found improvement in outcomes such as QoL, self-efficacy, and physical activity in the short term, no significant intervention effect was present on these outcomes when assessed at the 6-month follow-up. However, they noted that the intervention group demonstrated trends of improved levels of physical activity, whereas the control group did not. Reid et al [22] also reported long-term improvements in self-reported QoL and physical activity as long as 12 months from the start of the digital CR program, which was delivered over a 6-month period.

The examination of the studies by follow-up time revealed that the majority of the key outcome findings were mixed. However, all the studies reporting outcomes regarding adherence [19-21,26-30,35,37,38,40,41,43], QoL [15,17,21,22,24,25,27,29,33,35,38,39,41,42], and exercise capacity [15,17-22,24,25,27-33,35,38-42] found positive results or outcomes that were noninferior to the control group. Only studies with a follow-up period longer than 3 months reported outcomes for physical activity [14,15,22,25,31,32,36,41] and blood glucose control [29,31,35,41]. Regarding physical activity, the results were positive [14,15,22,25,31,32,36,41], whereas the results regarding blood glucose management were mixed, with positive effects [31], nonsignificant effects [29,35], or comparable results between the intervention and the control groups [41]. Similarly, the outcomes for blood pressure, depression or anxiety, and weight management were mixed (with positive effects, nonsignificant effects, or comparable results between the intervention and control groups) in both studies with shorter (3 months or less) [18,24,28,39] and longer (more than 3 months) [19,22,25,29-31,33,35,37,41,42] follow-up times. The only exceptions were as follows: one study reported a negative impact on mood at 6 months [16], 2 studies reported that the control group had better blood pressure management than the intervention group [25,35], and one study found an increase in body weight in the intervention group at 3 months [17]. For studies reporting outcomes regarding dietary habits, lipid management, and self-efficacy, studies with longer than a 3-month follow-up period reported positive or nonsignificant effects [15,16,19,25,29-32,35,37,41] or outcomes that were comparable between the intervention and the control groups [35,41], whereas studies with a follow-up period of 3 months or less found positive results [13,18,24]. The outcomes for nicotine dependence were mixed (positive or nonsignificant effects) among the studies with a short follow-up period [18,24,39], whereas one study reported no impact on smoking at 6 months [19].

The examination of more comprehensive, standalone interventions revealed that only 6 studies included in this review were standalone interventions delivering 5 or more CR components (other than a baseline assessment) [21,23-25,29,32]. The results of these studies for the majority of the key outcomes related to the CR components delivered through the interventions were heterogeneous. Of these 6 more comprehensive standalone interventions, only Brough et al [24] reported the outcomes regarding nicotine dependence, finding no impact on smoking. Devi et al [25] and Brough et al [24] found that their digital interventions had no significant impact on mood, whereas Varnfield et al [29] reported a positive effect on mood and anxiety levels in the intervention groups. Regarding dietary habits, Varnfield et al [29], Lear et al [32], and Brough et al [24] found improvement in the dietary habits of the intervention group, whereas Devi et al [25] found no significant impact. Three of these studies [24,25,29] reported an improvement in QoL in the intervention group, but Piotrowicz et al [21] found no significant difference in QoL measures between the intervention and control groups. Five of these studies reported a positive effect of digital intervention on exercise capacity [21,24,25,29,32]. In terms of weight management, Devi et al [25] and Varnfield et al [29] found a positive impact on weight in the intervention group, whereas Brough et al [24] found no impact on weight in the intervention group. Overall, there was a wide variety in the interventions delivered and outcomes reported.

Discussion

The Potential of Digital CR

This study highlights digital technology as a potential means of enhancing care and broadening access to CR through tailored interactive interventions. Our work differs from previous systematic reviews as our emphasis is on digital CR interventions with a focus on providing a systematic evaluation of the current literature to better understand the characteristics of these interventions. This study builds upon a growing body of literature supporting the use of internet-based features such as web portals and digital devices (eg, wearables) to remotely deliver CR components.

We found that digital CR was feasible and as effective as traditional CR in improving outcomes, whether as an adjunct or as an alternative to traditional CR [16,21,26,29,39,41,42]. Our findings support the conclusions of a previous study demonstrating a similar effectiveness of home- and center-based CR in improving clinical and health-related QoL outcomes in patients with myocardial infarction, myocardial revascularization, and heart failure [44]. In addition, Huang et al found that telehealth CR interventions were noninferior to center-based CR, both in the short term (12 weeks-1 year) and long term (up to 6 years), when comparing participants' exercise capacity, all-cause mortality, and modifiable risk factors, including blood pressure, blood lipids, smoking, and weight [45]. Moreover, the AACVPR, AHA, and ACC recently released a consensus statement highlighting evidence that home- and facility-based CR can achieve similar improvements in 3-12–month clinical outcomes [10]. These developments highlight digital technology as a potential means of enhancing care and broadening access to CR through tailored interactive interventions [46,47]. However, our study and literature indicate that until digital CR is further developed and better understood, there will be a need for in-person CR sessions (Multimedia Appendix 6). In-person sessions may help digital CR by establishing baseline and monitoring progress, personalizing

treatment plans, and bridging patient technology-usage challenges through technology education and deployment, especially for older users [48-50]. Currently, there are several ongoing clinical trials that are studying the efficacy of digital CR interventions [51-54].

Key Recommendations for Future Research

A limitation of our study is the heterogeneity of the identified papers, thus prohibiting meta-analysis. Papers varied in CR technologies, interventions, study design, measured outcomes, and control groups. For example, when considering a traditional CR population as a control group, some studies used standard care, some used direct comparison with facility-based CR, and other feasibility studies did not have a control comparison group. Although the diversity in studies proved challenging to quantify, it reiterates the motivation of this systematic review: digital health-based CR is emerging as an alternative or adjunct to standard CR; thus, methodologies have yet to reach a consensus. Consequently, we encourage practitioners to study digital CR approaches in a collaborative environment to promote the standardization and optimization of study methods. We have summarized our key recommendations in Textbox 3.

Textbox 3. Key recommendations for researchers conducting digital cardiac rehabilitation studies.

- Clearly state the specific goals of the intervention: whether it is designed to be a comprehensive standalone program or to be used adjunctively with traditional cardiac rehabilitation (CR)
- Describe the specifics of the CR components targeted through the intervention and the technology, equipment, or personnel required to deliver each of the components
- Include the specific details of the comparison group (eg, elaborate on what *usual care* consists of and include information about the specific intervention the comparison group received)

Beyond the heterogeneity of the study designs, we identified other limitations. Although we evaluated CR program components using the AHA and AACVPR's consensus statement on the core components of CR, the CR programs outside of the United States may differ from professional society guidelines in the United States. In addition, some studies included only minimal details about the specific components of their interventions; thus, the number of CR components delivered may have been greater than that captured in this review. Although studies reported the technology used (summarized in Multimedia Appendix 7), specific details were often sparse, limiting the evaluation of specifically what worked well and what was difficult to implement. Furthermore, the studies did not address long-term outcomes, as the maximum reported follow-up was 16 months, and most studies had fairly small sample sizes. Finally, the majority of the studies reviewed included sessions that were conducted in person for at least a portion of the overall CR intervention.

We identified several potential directions for future research. Although digital interventions have been found to successfully deliver the components of CR pertaining to physical activity, there remains a paucity of comprehensive digital CR program interventions that address risk factors such as lipid management, blood glucose level control, and smoking cessation. In addition, the long-term effectiveness of digital approaches to CR requires further evaluation [8,55]. Furthermore, studies with larger sample sizes and adequate control groups for comparison are necessary to better understand the impact of these interventions. Additional studies are also required to investigate the frequency of adverse events in patients participating in these interventions compared with traditional approaches. The adverse effects reported in the studies reviewed fell into the categories of (1) cardiac related (acute coronary syndrome, stable angina, arrhythmias, pericarditis, dyspnea, syncope, etc), (2) potential cardiac etiology (pleural effusions, cerebral ischemia, etc), (3) noncardiac related (peripheral artery disease, pneumonia, accidents, etc), and (4) death. A number of studies reported no significant difference in the rate of adverse effects between the digital intervention and control groups [14,30,33,38]. There were also studies that found that the intervention group had fewer associated adverse events than the control group; however, this did not always reach statistical significance [19,22,31,32,34]. Only one study reported that the digital CR intervention group had more adverse events during the treatment period as compared with the control group that participated in a center-based CR; however, adverse events were comparable during the postintervention follow-up period [41]. In some cases, no major adverse events occurred throughout the study [17,26-28,40,42], or the adverse effects were unrelated to the study intervention [16,21,35]. No conclusions could be drawn regarding the increased or decreased risk of adverse effects in some studies [15,18,20]. Overall, our findings highlight the need for robust digital intervention study designs with more comprehensive programs and analysis of both the short- and long-term effects.

Furthermore, as briefly mentioned in the Methods section, our team consulted with an external expert to include relevant original research on digital approaches to CR that may have been inadvertently excluded from our primary search strategy. We opted for this approach given that although the field of telehealth and telemedicine has grown rapidly over the past few decades, the adoption of common terminology remains in infancy [7]. The evolution of this terminology has been demonstrated in a bibliometric analysis by Fatehi and Wootton [6], who noted the emergence of terms such as *eHealth* and *mHealth* as well as the usage trends of the terms such as *telemedicine*, *telehealth*, and *eHealth* in the literature. As a result, there may be studies that include digital components, such as the usage of the internet, which our search terms failed to capture. Our experience further highlights the importance of more standardized terminology surrounding digital health interventions and an understanding of the evolving terminology to accurately review the existing literature.

Conclusions

Overall, we found that digital technology offers the potential to address the challenges associated with traditional, facility-based CR. If implemented on a large scale, digital CR could provide a level of impact, accessibility, affordability, cost savings, and benefits to patients not possible with conventional CR. However, so far, interventions have primarily focused on physical activity counseling and exercise training and not on the other core components of CR. In addition, our study focused on the evaluation of the technology used in digital CR and the comprehensiveness of these programs, but the intricacies of accreditation for CR programs are beyond the scope of this review. Further research is required with more comprehensive CR interventions to understand the long-term clinical impact of digital CR solutions on key cardiovascular outcomes and establish best practices for the development, delivery, and assessment of digital CR.

Acknowledgments

The authors thank Julie Nanavati, MLS, MA, (Johns Hopkins Welch Medical Library Informationist) for her assistance in developing the search strategy and downloading the references.

SW is supported by the Johns Hopkins School of Medicine Medical Scientist Training Program (National Institutes of Health: Institutional Predoctoral Training Grant – 5T32GM007309) and the National Institutes of Health: Ruth L. Kirschstein Individual Predoctoral NRSA for MD/PhD: F30 Training Grant (F30HL142131). ES is supported by the National Institutes of Health (NIH)/National Institute of Nursing Research (NINR) F31 NR017328, Ruth L. Kirschstein National Research Service Award and NIH/NINR T32 NR012704, predoctoral fellowship in interdisciplinary cardiovascular health research. KR was supported by grant TL1 TR003100 from the National Center for Advancing Translational Sciences, NIH. Funding sources for this project had no role in the design or conduct of the study or in data collection, management, or analysis, manuscript preparation, or in decisions about submitting the manuscript for publication. SM receives research support from the Aetna Foundation, the American Heart Association (20SFRN35380046 and COVID19-811000), the David and June Trone Family Foundation, the Pollin Digital Innovation Fund, Google, Patient-Centered Outcomes Research Institute (ME-2019C1-15328), the National Institutes of Health (P01 HL108800), Nokia, and the PJ Schafer Memorial Fund.

Authors' Contributions

SW, FM, and SM contributed substantially to the conception and design of this work. SW, RS, KR, and JW screened the references for inclusion in the review. NG provided external expert review on article selection. SW, EH, PH, HX, RS, KR, JW, and LS performed data extraction and analysis. YG and NO reviewed the studies to determine study quality. SW, EH, PH, HX, RS, KR, JW, YG, NO, LS, ES, GH, KK, WY, FM, EL, DM, NG, and SM provided a critical review of the manuscript.

Conflicts of Interest

FM and SM are founders of and hold equity in Corrie Health, which intends to further develop the digital platform. This arrangement has been reviewed and approved by the Johns Hopkins University in accordance with its conflicts of interest policies. Outside of the present work, they have received material support from Apple and iHealth and funding from the Maryland Innovation Initiative, Wallace H. Coulter Translational Research Partnership, Louis B. Thalheimer Fund, the Johns Hopkins Individualized Health Initiative, and the American Heart Association. SM reports personal fees for serving on scientific advisory boards for Akcea Therapeutics, Amgen, AstraZeneca, DalCor Pharmaceuticals, Esperion, Novo Nordisk, Quest Diagnostics, Regeneron, Sanofi, and 89bio. SM is a coinventor with a pending patent filed by the Johns Hopkins University for a system of low-density lipoprotein cholesterol estimation. EL is a scientific advisor for Moving Analytics. NG is the managing member of a population health management company, INTERVENT International, LLC.

Multimedia Appendix 1
Search Terms.
[DOCX File , 15 KB - jmir_v23i2e18773_app1.docx]

Multimedia Appendix 2
National Institutes of Health quality assessment tools: quality assessment of controlled intervention studies.
[DOCX File , 15 KB - jmir_v23i2e18773_app2.docx]

Multimedia Appendix 3
National Institutes of Health quality assessment tools: quality assessment tool for observational cohort and cross-sectional studies.
[DOCX File , 15 KB - jmir_v23i2e18773_app3.docx]

Multimedia Appendix 4

Summary of study and patient characteristics.
[DOCX File , 24 KB - jmir_v23i2e18773_app4.docx]

Multimedia Appendix 5
National Institutes of Health quality assessment for studies reviewed.
[DOCX File , 41 KB - jmir_v23i2e18773_app5.docx]

Multimedia Appendix 6
Aspects of in-person sessions.
[DOCX File , 19 KB - jmir_v23i2e18773_app6.docx]

Multimedia Appendix 7
Technology used in studies.
[DOCX File , 18 KB - jmir_v23i2e18773_app7.docx]

References

1. Cardiovascular disease. WHO. 2019. URL: http://www.who.int/cardiovascular_diseases/about_cvd/en/ [accessed 2020-12-29]
2. Anderson L, Oldridge N, Thompson DR, Zwisler A, Rees K, Martin N, et al. Exercise-based cardiac rehabilitation for coronary heart disease: Cochrane systematic review and meta-analysis. J Am Coll Cardiol 2016 Jan 05;67(1):1-12 [FREE Full text] [doi: 10.1016/j.jacc.2015.10.044] [Medline: 26764059]
3. Balady GJ, Williams MA, Ades PA, Bittner V, Comoss P, Foody JM, American Heart Association Exercise, Cardiac Rehabilitation,Prevention Committee, the Council on Clinical Cardiology, American Heart Association Council on Cardiovascular Nursing, American Heart Association Council on Epidemiology Prevention, American Heart Association Council on Nutrition, Physical Activity, Metabolism, American Association of Cardiovascular Pulmonary Rehabilitation. Core components of cardiac rehabilitation/secondary prevention programs: 2007 update: a scientific statement from the American Heart Association Exercise, Cardiac Rehabilitation, and Prevention Committee, the Council on Clinical Cardiology; the Councils on Cardiovascular Nursing, Epidemiology and Prevention, and Nutrition, Physical Activity, and Metabolism; and the American Association of Cardiovascular and Pulmonary Rehabilitation. Circulation 2007 May 22;115(20):2675-2682. [doi: 10.1161/CIRCULATIONAHA.106.180945] [Medline: 17513578]
4. Boyden T, Rubenfire M, Franklin B. Will increasing referral to cardiac rehabilitation improve participation? Prev Cardiol 2010;13(4):192-201 [FREE Full text] [doi: 10.1111/j.1751-7141.2010.00086.x] [Medline: 20879184]
5. Balady GJ, Ades PA, Bittner VA, Franklin BA, Gordon NF, Thomas RJ, American Heart Association Science Advisory Coordinating Committee. Referral, enrollment, and delivery of cardiac rehabilitation/secondary prevention programs at clinical centers and beyond: a presidential advisory from the American Heart Association. Circulation 2011 Dec 20;124(25):2951-2960. [doi: 10.1161/CIR.0b013e31823b21e2] [Medline: 22082676]
6. Fatehi F, Wootton R. Telemedicine, telehealth or e-health? A bibliometric analysis of the trends in the use of these terms. J Telemed Telecare 2012 Dec;18(8):460-464. [doi: 10.1258/jtt.2012.gth108] [Medline: 23209265]
7. Scott RE, Mars M. The same language speak we do - consensus terminology for telehealth. Stud Health Technol Inform 2016;231:99-109. [Medline: 27782021]
8. Bhavnani SP, Parakh K, Atreja A, Druz R, Graham GN, Hayek SS, et al. 2017 roadmap for innovation-ACC health policy statement on healthcare transformation in the era of digital health, big data, and precision health: a report of the American College of Cardiology task force on health policy statements and systems of care. J Am Coll Cardiol 2017 Nov 28;70(21):2696-2718 [FREE Full text] [doi: 10.1016/j.jacc.2017.10.018] [Medline: 29169478]
9. Sharma A, Harrington RA, McClellan MB, Turakhia MP, Eapen ZJ, Steinhubl S, et al. Using digital health technology to better generate evidence and deliver evidence-based care. J Am Coll Cardiol 2018 Jun 12;71(23):2680-2690 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.523] [Medline: 29880129]
10. Thomas RJ, Beatty AL, Beckie TM, Brewer LC, Brown TM, Forman DE, et al. Home-based cardiac rehabilitation: a scientific statement from the American Association of Cardiovascular and Pulmonary Rehabilitation, the American Heart Association, and the American College of Cardiology. J Am Coll Cardiol 2019 Jul 09;74(1):133-153 [FREE Full text] [doi: 10.1016/j.jacc.2019.03.008] [Medline: 31097258]
11. Study Quality Assessment Tools. National Heart, Lung, and Blood Institute. -. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2020-12-29]
12. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
13. Jenny NYY, Fai TS. Evaluating the effectiveness of an interactive multimedia computer-based patient education program in cardiac rehabilitation. The Occupational Therapy Journal of Research 2001 Aug 24;21(4):260-275. [doi: 10.1177/153944920102100404]

14. Barnason S, Zimmerman L, Nieveen J, Schulz P, Miller C, Hertzog M, et al. Influence of a symptom management telehealth intervention on older adults' early recovery outcomes after coronary artery bypass surgery. Heart Lung 2009;38(5):364-376 [FREE Full text] [doi: 10.1016/j.hrtlng.2009.01.005] [Medline: 19755186]
15. Maddison R, Pfaeffli L, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mobile phone intervention increases physical activity in people with cardiovascular disease: results from the HEART randomized controlled trial. Eur J Prev Cardiol 2015 Jun;22(6):701-709. [doi: 10.1177/2047487314535076] [Medline: 24817694]
16. Pfaeffli Dale L, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and internet support for coronary heart disease self-management: results from the text4heart randomized controlled trial. J Med Internet Res 2015 Oct 21;17(10):e237 [FREE Full text] [doi: 10.2196/jmir.4944] [Medline: 26490012]
17. Ades PA, Pashkow FJ, Fletcher G, Pina IL, Zohman LR, Nestor JR. A controlled trial of cardiac rehabilitation in the home setting using electrocardiographic and voice transtelephonic monitoring. Am Heart J 2000 Mar;139(3):543-548. [doi: 10.1016/s0002-8703(00)90100-5] [Medline: 10689271]
18. Gordon NF, English CD, Contractor AS, Salmon RD, Leighton RF, Franklin BA, et al. Effectiveness of three models for comprehensive cardiovascular disease risk reduction. Am J Cardiol 2002 Jun 01;89(11):1263-1268. [doi: 10.1016/s0002-9149(02)02323-8] [Medline: 12031725]
19. Southard BH, Southard DR, Nuckolls J. Clinical trial of an Internet-based case management system for secondary prevention of heart disease. J Cardiopulm Rehabil 2003;23(5):341-348. [doi: 10.1097/00008483-200309000-00003] [Medline: 14512778]
20. Scalvini S, Zanelli E, Comini L, Tomba MD, Troise G, Giordano A. Home-based exercise rehabilitation with telemedicine following cardiac surgery. J Telemed Telecare 2009;15(6):297-301. [doi: 10.1258/jtt.2009.090208] [Medline: 19720767]
21. Piotrowicz E, Baranowski R, Bilinska M, Stepnowska M, Piotrowska M, Wójcik A, et al. A new model of home-based telemonitored cardiac rehabilitation in patients with heart failure: effectiveness, quality of life, and adherence. Eur J Heart Fail 2010 Feb;12(2):164-171 [FREE Full text] [doi: 10.1093/eurjhf/hfp181] [Medline: 20042423]
22. Reid RD, Morrin LI, Beaton LJ, Papadakis S, Kocourek J, McDonnell L, et al. Randomized trial of an internet-based computer-tailored expert system for physical activity in patients with heart disease. Eur J Prev Cardiol 2012 Dec;19(6):1357-1354. [doi: 10.1177/1741826711422988] [Medline: 21903744]
23. Clark RA, Tideman P, Tirimacco R, Wanguhu K, Poulsen V, Simpson P, et al. A pilot study of the feasibility of an internet-based electronic Outpatient Cardiac Rehabilitation (eOCR) program in rural primary care. Heart Lung Circ 2013 May;22(5):352-359. [doi: 10.1016/j.hlc.2012.11.005] [Medline: 23294762]
24. Brough C, Boyce S, Houchen-Wolloff L, Sewell L, Singh S. Evaluating the interactive web-based program, activate your heart, for cardiac rehabilitation patients: a pilot study. J Med Internet Res 2014 Oct 29;16(10):e242 [FREE Full text] [doi: 10.2196/jmir.3027] [Medline: 25359204]
25. Devi R, Powell J, Singh S. A web-based program improves physical activity outcomes in a primary care angina population: randomized controlled trial. J Med Internet Res 2014 Sep 12;16(9):e186 [FREE Full text] [doi: 10.2196/jmir.3340] [Medline: 25217464]
26. Forman DE, LaFond K, Panch T, Allsup K, Manning K, Sattelmair J. Utility and efficacy of a smartphone application to enhance the learning and behavior goals of traditional cardiac rehabilitation: a feasibility study. J Cardiopulm Rehabil Prev 2014;34(5):327-334. [doi: 10.1097/HCR.0000000000000058] [Medline: 24866355]
27. Kraal JJ, Peek N, Van den Akker-Van Marle ME, Kemps HM. Effects of home-based training with telemonitoring guidance in low to moderate risk patients entering cardiac rehabilitation: short-term results of the FIT@Home study. Eur J Prev Cardiol 2014 Nov;21(2 Suppl):26-31 [FREE Full text] [doi: 10.1177/2047487314552606] [Medline: 25354951]
28. Piotrowicz E, Korzeniowska-Kubacka I, Chrapowicka A, Wolszakiewicz J, Dobraszkiewicz-Wasilewska B, Batogowski M, et al. Feasibility of home-based cardiac telerehabilitation: results of TeleInterMed study. Cardiol J 2014;21(5):539-536. [doi: 10.5603/CJ.a2014.0005] [Medline: 24526507]
29. Varnfield M, Karunanithi M, Lee C, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
30. Whittaker F, Wade V. The costs and benefits of technology-enabled, home-based cardiac rehabilitation measured in a randomised controlled trial. J Telemed Telecare 2014 Oct;20(7):419-422. [doi: 10.1177/1357633X14552376] [Medline: 25400004]
31. Frederix I, Van Driessche N, Hansen D, Berger J, Bonne K, Alders T, et al. Increasing the medium-term clinical benefits of hospital-based cardiac rehabilitation by physical activity telemonitoring in coronary artery disease patients. Eur J Prev Cardiol 2015 Feb;22(2):150-158. [doi: 10.1177/2047487313514018] [Medline: 24249840]
32. Lear SA, Singer J, Banner-Lukaris D, Horvat D, Park JE, Bates J, et al. Improving access to cardiac rehabilitation using the internet: a randomized trial. Stud Health Technol Inform 2015;209:58-66. [Medline: 25980706]
33. Smolis-Bąk E, Dąbrowski R, Piotrowicz E, Chwyczko T, Dobraszkiewicz-Wasilewska B, Kowalik I, et al. Hospital-based and telemonitoring guided home-based training programs: effects on exercise tolerance and quality of life in patients with heart failure (NYHA class III) and cardiac resynchronization therapy. A randomized, prospective observation. Int J Cardiol 2015 Nov 15;199:442-447. [doi: 10.1016/j.ijcard.2015.07.041] [Medline: 26276068]

34. Frederix I, Hansen D, Coninx K, Vandervoort P, Vandijck D, Hens N, et al. Effect of comprehensive cardiac telerehabilitation on one-year cardiovascular rehospitalization rate, medical costs and quality of life: a cost-effectiveness analysis. Eur J Prev Cardiol 2016 May;23(7):674-682. [doi: 10.1177/2047487315602257] [Medline: 26289723]
35. Skobel E, Knackstedt C, Martinez-Romero A, Salvi D, Vera-Munoz C, Napp A, et al. Internet-based training of coronary artery patients: the Heart Cycle Trial. Heart Vessels 2017 Apr;32(4):408-418. [doi: 10.1007/s00380-016-0897-8] [Medline: 27730298]
36. Thorup C, Hansen J, Grønkjær M, Andreasen JJ, Nielsen G, Sørensen EE, et al. Cardiac patients' walking activity determined by a step counter in cardiac telerehabilitation: data from the intervention arm of a randomized controlled trial. J Med Internet Res 2016 Apr 04;18(4):e69 [FREE Full text] [doi: 10.2196/jmir.5191] [Medline: 27044310]
37. da Silva Vieira AS, Melo C, Machado J, Gabriel J. The effect of virtual reality on a home-based cardiac rehabilitation program on body composition, lipid profile and eating patterns: a randomized controlled trial. Eur J Prev Cardiol 2017 Jan;9:69-78. [doi: 10.1016/j.eujim.2016.11.008]
38. Hwang R, Bruning J, Morris NR, Mandrusiak A, Russell T. Home-based telerehabilitation is not inferior to a centre-based program in patients with chronic heart failure: a randomised trial. J Physiother 2017 Apr;63(2):101-107 [FREE Full text] [doi: 10.1016/j.jphys.2017.02.017] [Medline: 28336297]
39. Fang J, Huang B, Xu D, Li J, Au WW. Innovative application of a home-based and remote sensing cardiac rehabilitation protocol in Chinese patients after percutaneous coronary intervention. Telemed J E Health 2019 Apr;25(4):288-293. [doi: 10.1089/tmj.2018.0064] [Medline: 30192210]
40. Harzand A, Witbrodt B, Davis-Watts ML, Alrohaibani A, Goese D, Wenger NK, et al. Feasibility of a smartphone-enabled cardiac rehabilitation program in male veterans with previous clinical evidence of coronary heart disease. Am J Cardiol 2018 Nov 01;122(9):1471-1476 [FREE Full text] [doi: 10.1016/j.amjcard.2018.07.028] [Medline: 30217377]
41. Maddison R, Rawstorn JC, Stewart RAH, Benatar J, Whittaker R, Rolleston A, et al. Effects and costs of real-time cardiac telerehabilitation: randomised controlled non-inferiority trial. Heart 2019 Jan;105(2):122-129 [FREE Full text] [doi: 10.1136/heartjnl-2018-313189] [Medline: 30150328]
42. Peng X, Su Y, Hu Z, Sun X, Li X, Dolansky MA, et al. Home-based telehealth exercise training program in Chinese patients with heart failure: A randomized controlled trial. Medicine (Baltimore) 2018 Aug;97(35):e12069 [FREE Full text] [doi: 10.1097/MD.0000000000012069] [Medline: 30170422]
43. Rawstorn JC, Gant N, Rolleston A, Whittaker R, Stewart R, Benatar J, et al. End users want alternative intervention delivery models: usability and acceptability of the remote-cr exercise-based cardiac telerehabilitation program. Arch Phys Med Rehabil 2018 Nov;99(11):2373-2377. [doi: 10.1016/j.apmr.2018.06.027] [Medline: 30076800]
44. Anderson L, Sharp GA, Norton RJ, Dalal H, Dean SG, Jolly K, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2017 Jun 30;6:CD007130 [FREE Full text] [doi: 10.1002/14651858.CD007130.pub4] [Medline: 28665511]
45. Huang K, Liu W, He D, Huang B, Xiao D, Peng Y, et al. Telehealth interventions versus center-based cardiac rehabilitation of coronary artery disease: a systematic review and meta-analysis. Eur J Prev Cardiol 2015 Aug;22(8):959-971. [doi: 10.1177/2047487314561168] [Medline: 25488550]
46. Rawstorn JC, Gant N, Direito A, Beckmann C, Maddison R. Telehealth exercise-based cardiac rehabilitation: a systematic review and meta-analysis. Heart 2016 Aug 01;102(15):1183-1192. [doi: 10.1136/heartjnl-2015-308966] [Medline: 26936337]
47. Su JJ, Yu DSF, Paguio JT. Effect of eHealth cardiac rehabilitation on health outcomes of coronary heart disease patients: a systematic review and meta-analysis. J Adv Nurs 2020 Mar;76(3):754-772. [doi: 10.1111/jan.14272] [Medline: 31769527]
48. Emerging technologies to support an aging population. TASK FORCE ON RESEARCH AND DEVELOPMENT FOR TECHNOLOGY TO SUPPORT AGING ADULTS 2019:1-40 [FREE Full text]
49. Andrews JA, Brown LJ, Hawley MS, Astell AJ. Older adults' perspectives on using digital technology to maintain good mental health: interactive group study. J Med Internet Res 2019 Feb 13;21(2):e11694 [FREE Full text] [doi: 10.2196/11694] [Medline: 30758292]
50. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
51. Duscha BD, Piner LW, Patel MP, Craig KP, Brady M, McGarrah RW, et al. Effects of a 12-week mHealth program on peak VO and physical activity patterns after completing cardiac rehabilitation: a randomized controlled trial. Am Heart J 2018 May;199:105-114. [doi: 10.1016/j.ahj.2018.02.001] [Medline: 29754647]
52. Identifier: N. Longterm follow-up of cardiac patients with an smartphone-application -a randomized clinical study. - 2017:1.
53. Medical Research C. Creating an Optimized Technology-Based Weight Loss Program for Cardiac Rehabilitation Patients. CMR 2000 Jan 01:1. [doi: 10.31525/ct1-nct03845283]
54. Networks Inc. K. Rehabilitation at home using mobile health in older adults after hospitalization for Ischemic Heart Disease (resilient). Study Record Detail 2019 Jun 06:-. [doi: 10.31525/ct1-nct03978130]
55. Salvi D, Ottaviano M, Muuraiskangas S, Martínez-Romero A, Vera-Muñoz C, Triantafyllidis A, et al. An m-Health system for education and motivation in cardiac rehabilitation: the experience of HeartCycle guided exercise. J Telemed Telecare 2018 May;24(4):303-316. [doi: 10.1177/1357633X17697501] [Medline: 28350282]

Abbreviations

AACVPR: American Association of Cardiovascular and Pulmonary Rehabilitation
ACC: American College of Cardiology
AHA: American Heart Association
CR: cardiac rehabilitation
CVD: cardiovascular disease
mHealth: mobile health
NHLBI: National Heart, Lung, and Blood Institute
NIH: National Institutes of Health
NINR: National Institute of Nursing Research
QoL: quality of life

Edited by G Eysenbach; submitted 17.03.20; peer-reviewed by L Quinlan, A Triantafyllidis, F Holl, M Mars, C Carrion; comments to author 29.06.20; revised version received 22.11.20; accepted 07.12.20; published 08.02.21.

Please cite as:
Wongvibulsin S, Habeos EE, Huynh PP, Xun H, Shan R, Porosnicu Rodriguez KA, Wang J, Gandapur YK, Osuji N, Shah LM, Spaulding EM, Hung G, Knowles K, Yang WE, Marvel FA, Levin E, Maron DJ, Gordon NF, Martin SS
Digital Health Interventions for Cardiac Rehabilitation: Systematic Literature Review
J Med Internet Res 2021;23(2):e18773
URL: https://www.jmir.org/2021/2/e18773
doi:10.2196/18773
PMID:33555259

Review

Online Communities as a Driver for Patient Empowerment: Systematic Review

Victoria Johansson[1*], MSc; Anna Sigridur Islind[1,2*], PhD; Tomas Lindroth[1,3*], PhD; Eva Angenete[4,5*], PhD; Martin Gellerstedt[1,6*], PhD

[1]University West, School of Business, Economics and IT, SE-461 86, Trollhättan, Sweden

[2]School of Computer Science, Reykjavik University, Reykjavik, Iceland

[3]Department of Applied IT, University of Gothenburg, Gothenburg, Sweden

[4]Department of Surgery, SSORG - Scandinavian Surgical Outcomes Research Group, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[5]Region Västra Götaland, Sahlgrenska University Hospital/Östra, Department of Surgery, Gothenburg, Sweden

[6]School of Health Sciences, University of Skövde, Skövde, Sweden

[*]all authors contributed equally

Corresponding Author:
Victoria Johansson, MSc
University West
School of Business, Economics and IT
Gustava Melins gata 2
461 86 Trollhättan
Sweden
Phone: +46760503052
Email: victoria.johansson@hv.se

Abstract

Background: The use of online resources has changed how people manage health care processes. Patients seek information about health conditions, guidance in treatment, and support from peers online, complementary to traditional health care trajectories. Online communities have the potential to contribute to the quality of care by increasing patient empowerment; however, there is a gap in research regarding in what way online communities contribute to patient empowerment.

Objective: We synthesized research regarding how online communities contribute to patient empowerment to address the research question “In what ways can participation in online communities support patient empowerment?” by studying how patient empowerment is operationalized in different studies. The definition of patient empowerment used in this paper is enablement for people to develop mastery over actions and control over decisions that influence their lives. The mastery is both through processes and outcomes of the development.

Methods: A systematic review was conducted by searching in the following databases: Scopus, ACM Digital Library, EBSCO (CINAHL and MEDLINE), PubMed, and Web of Science. In total, there were 1187 papers after excluding duplicates, and through selection processes using an analytical framework with definitions of patient empowerment and related concepts, 33 peer-reviewed papers were included.

Results: Findings indicated that online communities support patient empowerment both as a process and as outcomes of these processes. Additionally, it was seen as a complement to traditional health care and encouragement for health care professionals to have a more positive attitude toward patients’ usage. There was a mix between deductive (19/33, 58%), inductive (11/33, 33%), and a mixed approach (3/33, 9%) of studying patient empowerment in various forms. The online communities in most papers (21/33, 64%) were well-established and represented patients’ initiatives.

Conclusions: There is a need to include professionals' perspectives regarding how health care can embrace patient empowerment through online communities. This systematic review's main contribution is the proposal of a new framework and conceptualization of how patient empowerment in online communities can be understood from different hierarchical levels.

(J Med Internet Res 2021;23(2):e19910) doi:10.2196/19910

KEYWORDS

patient empowerment; online community; person-centered care; eHealth; systematic review

Introduction

Background

When a person faces a difficult situation, for example, when receiving a medical diagnosis, a fear of not being able to control the outcome—feeling disempowered—is a natural response [1]. Up until recently, health care professionals have been the primary resource for helping patients regain empowerment by finding suitable treatment and giving recommendations and support.

Potential of Online Communities and Patient Empowerment

In parallel to efforts provided by health care, now more than ever, patients use the internet and online communities as complementary health care resources. In a study of the US population, it was found that 80% of internet users (74% of the population) looked for health-related issues, and 18% had gone online to find peers with similar health concerns. This fundamentally changes how people manage their care process, as patients are able to seek guidance, experiences, and support from peers as complementary resources to manage condition of illness and potential posttreatment with the aim of returning to the new normal self [2]. A review [3] found that previous studies have shown that patients use online communities because they experience or believe that health care professionals filter information; are unaware of the latest research; and lack the capability of showing empathy. Thus, online communities serve as supporting resources to increase information and emotional support. Another argument is to get the first-hand experience as a complement to health care expertise, which may help patients translate recommendations and instructions into daily self-care strategies. This could expand patients' knowledge regarding health conditions and treatments while also helping them find emotional and social support [4], thereby increasing empowerment.

Patient empowerment refers to processes and outcomes at both individual and group level that enable people to develop mastery over actions and control over decisions that influence their lives [5-7]. Patient empowerment could be regarded as being complementary to person-centered care [8,9]. Person-centered care focuses on designing and delivering individualized care, while patient empowerment focuses on a modified relationship between patients and health care professionals that enables patient-driven and patient-centered care [10]. Research indicates that patient-centered approaches are usually more cost-effective [11-13]. Common ground in both concepts is more engaged and informed patients, that is, more empowered patients.

Difficulties in Online Community Research

Unarguably, the use of online communities has potential, given the right conditions, to be beneficial for patients. The aggregated knowledge found in online communities could serve as a tool for professionals and the whole health care system for quality improvement in health service delivery [14-16]. In this way, online communities could be a contributor to change in health care. However, evidence of efficacy is equivocal with varying results, and comprehensive reviews [3,17-19] show no evidence of harm, but no strong evidence of efficacy either. The same reviews [3,17-19] reported that many of their included studies had methodological weaknesses. Additionally, recent reviews [3,20] reported difficulties in making comparisons between studies due to methodological problems and lack of analytical frameworks [3,20]. Furthermore, benefits are frequently discussed in relation to individual patients. Other factors includes the various challenges posed by usability and sociability of online communities [21-23].

Objective

Since difficulties in comparing of methodology and efficacy have already been illustrated, this was not the aim of this review. A part of the methodological problem may be lack of a well-established definition of patient empowerment or comprehensive framework related to different levels of the empowerment concept on an individual and collective level in relation to online communities. Therefore, the primary objective for this systematic review was to clarify in which ways participation in online communities can support patient empowerment. This was done by studying how patient empowerment in online communities has been operationalized in different studies.

Methods

Information Sources and Eligibility Criteria

The structure of the review followed principles of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [24]. Specific principles that were followed in the manuscript are presented in Multimedia Appendix 1. Complementary resources that guided the structure were inspired by recently published systematic reviews in the Journal of Medical Internet Research [20,25-31].

For this systematic review, Scopus, ACM Digital Library, EBSCO (CINAHL and MEDLINE), PubMed, and Web of Science were searched. We defined inclusion and exclusion criteria that did not depend on time limitation or a particular research field to receive a high variety of papers and see potential differences or similarities regarding operationalization and definitions of empowerment (Multimedia Appendix 2). Additionally, this decision was made since both patient empowerment and online communities have been researched in a variety of research fields and used different notions. For instance, online community, a notion more recently used based on the phenomenon it refers to for people communicating in so-called internet forums, started in the late 1970s [32].

Search Strategy

Searches in all databases were performed on January 17, 2019 by the first 3 authors. In order to capture relevant papers, 2 search strings were used: (1) patient empowerment and related concepts, and (2) online communities and related concepts.

These 2 search strings were combined and used as main search strategy (Figure 1, Multimedia Appendix 3). These 2 search strings included words with a similar meaning or used interchangeably with patient empowerment and online communities. Through this search strategy, we found 1187 references after removing duplicates.

Figure 1. Main search strategy.

Patient Empowerment Keywords		Online Community Keywords
"patient empowerment" **OR**		"online communit*" **OR**
"patient activation" **OR**		"online peer-support" **OR**
"patient enablement" **OR**		"online interpersonal communication" **OR**
"patient engagement" **OR**	**AND**	"online health communit*" **OR**
"patient involvement" **OR**		"online patient support" **OR**
"patient participation"		"online social support" **OR**
		"online peer-to-peer support"

Construction of Search Strings

For the patient empowerment search string, we came to rely on the work of defining patient empowerment and related concepts by [5,6] since their work had the purpose of finding a consensus definition of patient empowerment based on previously published research, wherein patient empowerment is described as an umbrella concept—the related concepts (Figure 1) are part of what is considered to be the main definition that the concept entails. Therefore, we did a compilation framework of these definitions of patient empowerment through the related concepts, which was later used as an analytical framework (Multimedia Appendix 4). For the online community search string, all authors did brainstorming sessions together, and the search string was constructed in relation to the inclusion criteria of the Online Community Perspective (Table S1, Multimedia Appendix 2) and through pilot searches of keywords and suggestions of keywords in the selected databases. The combination of the 2, resulted in the final selection of keywords (Figure 1).

Selection Process

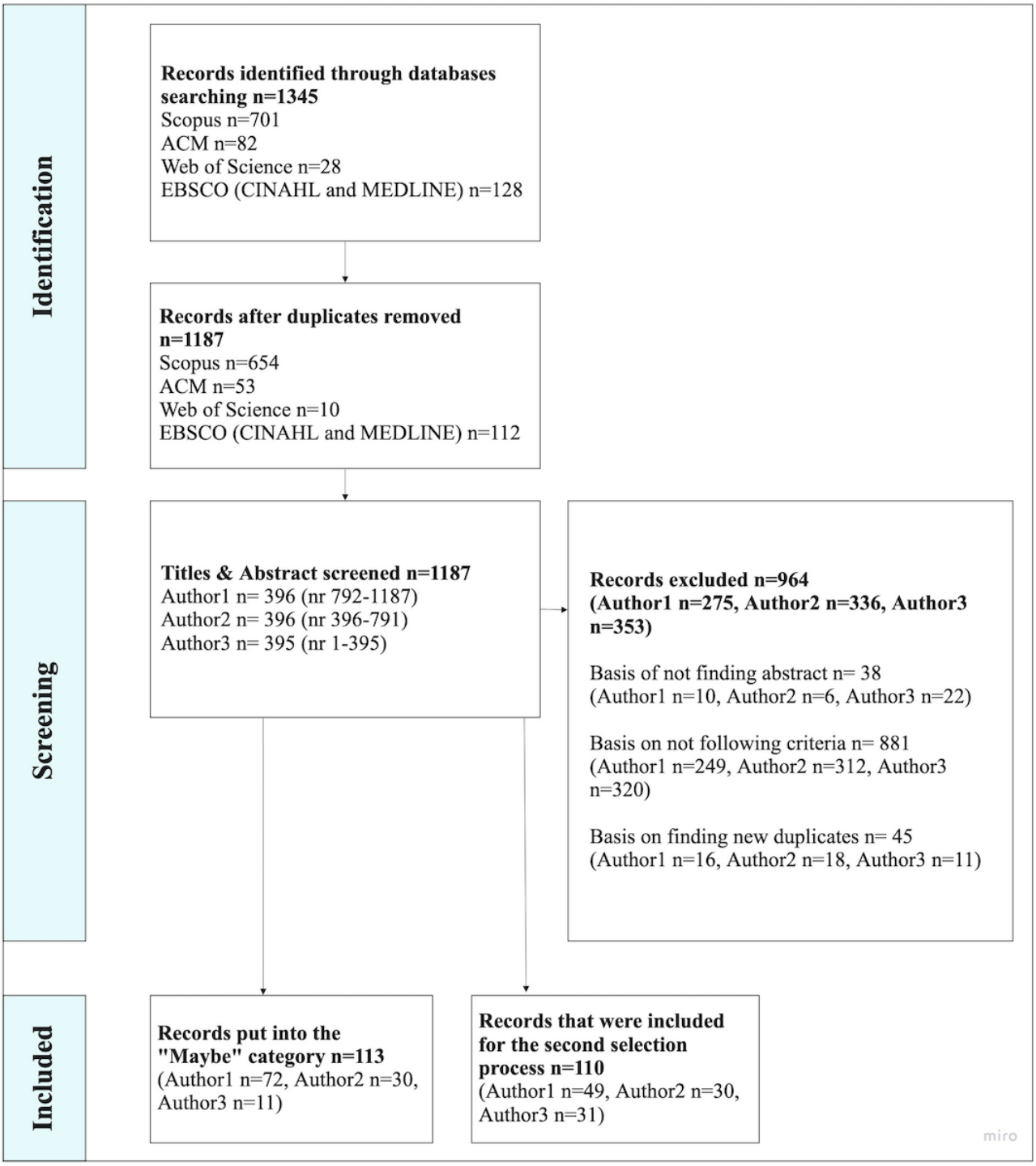
Overview

The selection process was conducted by the first 3 authors, in 2 phases. In phase 1, titles, abstracts, and keywords were screened. Phase 2 involved in-depth reading of full texts. The inclusion and exclusion criteria were applied in both phases (Table S1 and Table S2, Multimedia Appendix 2).

First Selection Process

After the first selection process, 223 papers were included (Figure 2). As seen in Figure 2, there was a category named *Maybe*, containing papers for which uncertainty existed regarding inclusion criteria based on only the content of the abstract. Therefore, these papers, alongside the papers with abstracts meeting the inclusion criteria from the first selection process moved forward to the second selection phase.

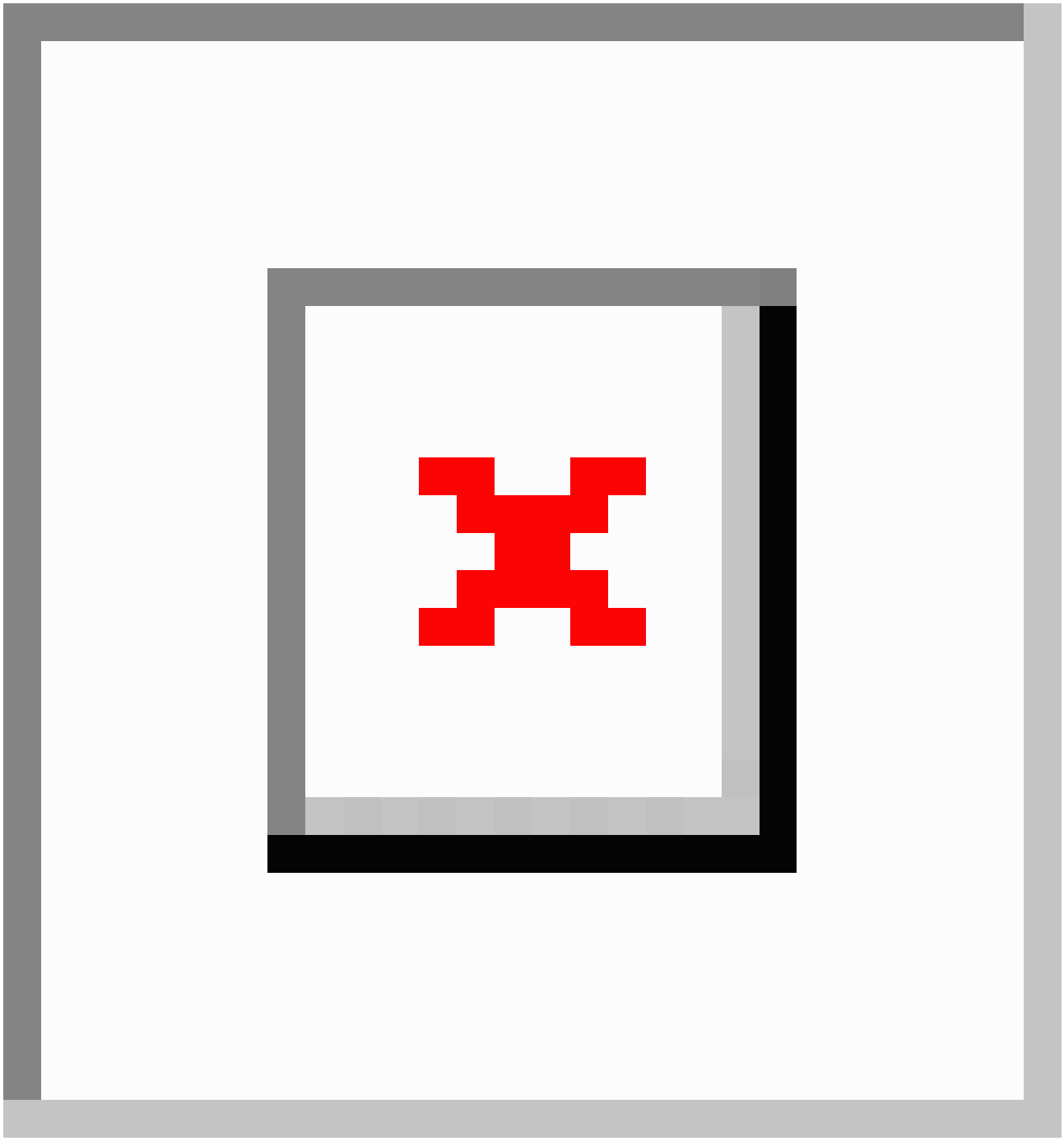
Figure 2. First selection process.



Second Selection Process

Papers included in the second phase were assigned to a different screener than the original screener to ensure intercoder reliability (Figure 3, Multimedia Appendix 5). This phase resulted in a final selection of 33 papers. Papers were excluded based on (1) not finding the full text; (2) not meeting the inclusion criteria; (3) in the process of conducting the synthesis of results; (4) meeting the exclusion criteria or it was unclear how, for instance, patient empowerment was evidenced. For the third exclusion reason, after having finalized the second process, we reread each paper multiple times to identify how each paper met the inclusion criteria. Therefore, we discovered that some of the papers that had been included met our exclusion criteria regarding how patient empowerment was evidenced, creating difficulty in remaining neutral to objectively see the characteristics of the paper, without enforcing our own interpretation of how it could answer the research question, and if included, would go against the purpose of conducting systematic literature reviews [33-35] (Table S2, Multimedia Appendix 2).

Figure 3. Second selection process.



Synthesis of Results and Analytical Framework

Two synthesis matrixes were designed prior to analysis to organize potential findings and identify factors involved in these findings [34,36]. The first involved themes of structure and format of each selected paper, that is, author, year, title, format of article, method, and type of online community. The second matrix contained analytical themes that were structured according to frameworks that contained definitions of patient empowerment developed by [5,6] (Multimedia Appendix 4). The analytical framework in the second matrix functioned as a guide to complete the final selection processes and was also used to guide the analysis of results. Each of the 33 papers included were thereby reported in both matrixes and completed with summaries. Finally, all authors were involved in discussion of the synthesis of the results.

Results

Overview

The results are divided into 2 parts:

1. Main characteristics: the systematic summarization of selected papers, presentation of main characteristics, that is, type and initiation of online communities, and approach of studying patient empowerment
2. How online communities support patient empowerment: a synthesis of findings in relation to the research question

Main Characteristics

Papers

The papers included in this systematic review were published between 2000 and 2018. Specific details regarding year of publication, journal, methodology, and other related characteristics are presented in Multimedia Appendix 6.

Type and Initiation of Online Communities

In the selection, 21 out of 33 papers (64%) presented an online community that was an established community [37-57]; in 7 out of 33 papers (21%), the community type was undefined [58-64], and in 5 papers (15%), the researchers had designed their own community [65-69].

Established and Undefined Online Community Papers

The established communities targeted specific diagnoses, were communities that were available 24/7, and had many visitors or members over time. The communities existed before being studied (Multimedia Appendix 6). In 2 of these studies, the community was presented as a built-in community through social media, for example, Facebook [40,47]. There were established communities that did not specify the platform, merely that they enabled peer support (3/33, 9%) [46,50,54], while others presented well-established online communities for a specific diagnosis (11/33, 33%) [37-39,41-43,52,54-57]. Additionally, the established online communities were based on patients' initiatives, both regarding creating (6/33, 18%) [40,43-45,52,57], maintaining and moderating (5/33, 15%) [40,43-45,52], or using the community (21/33, 64%) [37-57].

The papers that had undefined communities mainly aimed to achieve understanding of patient narratives and general experiences from using online communities, instead of the technology in particular. Thus, in 5 of the 7 papers (15%), the aim was to map different behaviors or reasons for use in order to understand the respondents' levels of patient empowerment [59,61-64], while in 3 out of 7 (9%), the aim was to understand which factors contributed to patient empowerment when using online communities [58,60,64].

Involvement of Health Care Professionals

The papers that represented either established or undefined online communities (28/33, 85%) did not discuss or analyze involvement of health care professionals. If health care professionals were mentioned, it was merely to explain why patients used the community (24/33, 73%) [37,39,40,43-62,64], if patients chose to share experiences of use (5/33, 15%) [45,48,49,58,59], or professionals contributed to content in various forms (5/33, 15%), but it was not elaborated on how or if this content had an effect on patients [45,49,53,54,56].

Designed Community and Involvement of Health Care Professionals

In the papers that reported having designed online communities, the format was either a web platform [65], online forum [66], own software [68], a list server that created email threads [69], or an e-recovery portal containing online community functionalities [67]. The main function of all designed communities was patient-peer forums. Other functionalities presented were password protection (3/33, 9%) [65,67,68], ability to individually contact health care professionals and design care plans (1/33, 3%) [67], or moderated discussion boards in various formats that was led by researchers (5/33, 15%) [65-69].

In 3 out of 5 papers (3/33, 9%), initiation of use was through joint consultation with health care professionals [65,67,68]. This meant that patients were recruited or recommended by their health care professional in order to participate in a specific online community and study. One of these 3 papers [67] had the health care professionals as part of the study's results.

Studying Patient Empowerment

Of 33 papers, 22 (70%) focused on studying patient empowerment explicitly (Table 1). The focus referred to either a deductive (12/33, 36%) [37-39,47,50-52,55,57,61,65,69] or inductive approach (13/33, 39%) [40,43,46-49,54,57-60,66,67], or a mix thereof. Papers with an inductive approach often revealed patient empowerment as conclusions of thematic analysis or through discussion of findings. Papers with a deductive approach relied on definitions, research, and measurement scales developed by, for example, van Uden-Kraan et al [38,39,70] (7/33, 21%) [37-39,50,51,55,57], Zimmerman [7] (2/33, 6%) [51,53], Spreitzer [71] (2/33, 6%) [47,65], and Barak et al [72] (3/33, 9%) [50,51,57].

The 10 remaining papers (10/33, 30%) did not use patient empowerment explicitly but studied related concepts that are either part of the related concepts presented in the analytical framework (Multimedia Appendix 4) or are related concepts, that is, not explicitly defined in the analytical framework (Table 1). These concepts are presented in Table 1. However, the common denominator of these papers was that they studied specific concepts in a deductive way, thus used relevant measurements and analytical frameworks from previous research of the concept in question.

Table 1. Papers that studied patient empowerment or other related concepts.

Characteristic	Established online community		Undefined online community		Designed online community	
	Papers, n (%)	References	Papers, n (%)	References	Papers, n (%)	References
Total papers (N=33)[a]	21 (64)	[37-57]	7 (21)	[58-64]	5 (15)	[65-69]
Concept studied						
Patient empowerment	15 (45)	[37-40,43,46-52,54,55,57]	4 (12)	[58-61]	4 (12)	[65-67,69]
Patient activation	1 (3)	[45]	0 (0)	N/A[b]	0 (0)	N/A
Patient engagement	0 (0)	N/A	0 (0)	N/A	1 (3)	[68]
Adherence to treatment	2 (6)	[44,56]	1 (3)	[64]	0 (0)	N/A
Self-reappraisal	0 (0)	N/A	1 (3)	[63]	0 (0)	N/A
Self-efficacy	1 (3)	[56]	0 (0)	N/A	0 (0)	N/A
Cyber-informational and decisional empowerment	0 (0)	N/A	1 (3)	[62]	0 (0)	N/A
Intrapersonal and interactional aspect of psychological empowerment	1 (3)	[53]	0 (0)	N/A	0 (0)	N/A
Individual and collective empowerment	0 (0)	N/A	1 (3)	[64]	0 (0)	N/A
Well-being	2 (6)	[41,42]	0 (0)	N/A	0 (0)	N/A
Emotional coping	1 (3)	[41]	0 (0)	N/A	0 (0)	N/A

[a]All percentages refer to the total number of papers, N=33.

[b]N/A: not applicable.

How Online Communities Support Patient Empowerment

Framework

This part will follow the analytical framework structure of presenting different concepts related to patient empowerment: patient enablement, activation, engagement, involvement, and participation (Multimedia Appendix 4).

Patient Enablement

Patient enablement is usually presented as the starting phase of becoming empowered and is defined as (1) the possibilities and prerequisites that health care gives the patient to self-manage their own health condition [5] or (2) the patient's confidence in the ability to improve management of condition or the relationship with health care professionals [6].

Established and Undefined Online Communities

For established and unspecified online communities, patient enablement was in 25 out of 33 papers (75%) analyzed as the prerequisite that an online community had for the patient to become engaged and activated in managing diagnosis [37-52,54,55,57-60,62-64]. The prerequisites were often related to patients' confidence in their own abilities to improve their health conditions or relationships with health care professionals. Patient confidence was analyzed as the context behind the use of online communities, which was divided into (1) becoming better informed, for example, about coping with different treatment alternatives, in order to become more involved in decision-making processes during consultation; (2) coping with the emotional burden of diagnosis in everyday life, by reading and writing content or networking with others with shared experiences; or (3) absorbing and reflecting on information that was perceived as missing or not fully elaborated on during health care consultations (Table 2).

Table 2. The direction of context behind patient confidence in selected references.

Characteristic	Established online community		Undefined online community		Designed online community	
	Papers, n (%)	References	Papers, n (%)	References	Papers, n (%)	References
Total papers (N=33)[a]	21 (64)	[37-57]	7 (21)	[58-64]	5 (15)	[65-69]
Context behind patient confidence						
To become better informed	17 (52)	[37-40,43-49,51,53-57]	4 (12)	[58,60,61,64]	4 (12)	[65-68]
Coping with emotional burden and networking with peers	17 (52)	[37,40-49,51,53-57]	4 (12)	[58,60,61,64]	5 (15)	[65-69]
Absorb and reflect on information given by health care	15 (45)	[37,40,43-49,51,53-57]	4 (12)	[58,60,61,64]	2 (6)	[65,67]

[a]All percentages refer to the total number of papers, N=33.

Designed Online Communities

Similar prerequisites were identified in all of the designed community papers (Table 2) since functionalities were tailored according to the type of intervention being studied; the directions were (1) developing patients' understanding of when to seek consultation (to become better informed); (2) providing inspiration through other patients' stories in order to boost self-confidence in self-management of health condition (coping with emotional burden and networking with peers); (3) being equally involved in decision-making processes (to become better informed and absorb and reflect on information given by health care). An additional intervention purpose listed in all 5 community papers was how to expand research or health care services further in order to adapt to patients' needs (5/33, 15%) [65-69].

The differences for the designed communities in comparison with the established communities were that the prerequisites were given by health care professionals or researchers, by informing patients during recruitment about the purpose of usage, which led to patients' participation in the study, thus potential engagement and activation of health through the online community. Hence, the health care professionals were the leading part in the patient-provider relationship, since the recruitment was determined by the professionals' confidence in the individual patient's way of improving self-management through usage of community. However, during later phases of use, it was the patient who had the confidence in their ability, based on inspiration from other patients' support in a particular online community.

Patient Activation and Engagement

Patient activation is described as the phase when patients act through knowledge gained, and create intermediate goals in order to improve their health condition. Additionally, it is presented as the phase in which patients know where to acquire knowledge and what they need to do to receive it [5,6]. Patient activation is often considered to be intertwined with patient engagement, which is defined as the patient's motivation for improving health condition through a collaborative relationship with health care professionals. To motivate patients, health care professionals need to make patients aware of care processes, which is thereby the first step required to create good conditions for patient involvement and participation [5,6].

Established and Undefined Online Communities

In the established and unspecified online community papers, patient activation and patient engagement were analyzed as an integrated process that was supported by patient peers and not by health care professionals. This integrated process was shown by the use of the online communities—how patients searched and absorbed needed information. Thus, this generated how open patients were to (1) change health habits, (2) wanting to help others in the online community, or (3) wanting to prepare for upcoming care consultation with health care professionals (Table 3). The context behind what was generated as motivation is presented in Table 3.

Table 3. The integrated processes and context of motivation in established and undefined community papers (N=28/33).

Characteristic	Established online community		Undefined online community	
	Papers, n (%)[a]	References	Papers, n (%)[a]	References
Total papers (n=28)	21 (64)	[37-57]	7 (21)	[58-64]
The integrated process				
Change health habits	8 (24)	[37,40,45-47,51,56,57]	5 (15)	[58-60,63,64]
Helping others	14 (42)	[38-47,50,51,55,57]	4 (12)	[58-60,64]
Prepare for upcoming care consultation with health care professionals	8 (24)	[39,40,44,46,49,52,56,57]	5 (15)	[46,58,59,61,62]
Context behind motivation				
Take control over health	19 (58)	[37-40,43-57]	7 (21)	[58-64]
Improve ability and conditions for patient involvement	10 (30)	[37,40,43,44,46-49,54,56]	5 (15)	[59-62,64]
Heal emotionally	14 (42)	[38-47,50,51,55,57]	4 (12)	[58-60,64]

[a]All percentages refer to the total number of papers, N=33.

Designed Online Communities

In the designed online community papers, patient activation and patient engagement were analyzed as separate processes. The role of online community support for patient activation was to create independence and was presented as patients taking part in other patients' narratives and discussion with patient peers. In 2 out of the 5 papers (2/33, 6%), it was described as reasons for patients to be able to construct individual goals that were relevant to the individual situation and presented as a basis for the decision-making process during health consultations [65,67]. Identification of patient engagement within 3 of 5 designed community papers (3/33, 9%) was analyzed as patients' motivation to create good conditions for collaboration with health care professionals to understand how the collaboration could generate better health outcomes [65,67,68].

Additional Measurements of Patient Engagement

Patient engagement was also studied by and presented as measuring how active patients were in an online community in the form of number of visits, time spent on the online community, or whether patients had contributed to content or not. These measures were used in one designed community paper (1/33, 3%) [68] and in 5 established community papers (5/33, 15%). The established community papers did not focus on studying patient engagement but used patient engagement measures such as those previously mentioned, in order to study their selected concept of patient empowerment (Table 1) [38,39,41,42,53].

Patient Involvement and Participation

Patient involvement is presented as (1) an advanced phase of patient engagement through patients' awareness of the patient role within different care processes, which thus contributes to a collaborative relationship with health care professionals [6], or (2) health care providers' prerequisites to include the patient during consultation as a first step for a collaborative relationship, which will later lead to the patient being the one who determines the prerequisites for consultation and decision making. This latter phase is presented as *patient participation* by [5] and *patient involvement* by [6].

Established and Undefined Online Communities

The analysis of patient involvement and participation was identified through the outcome of using online communities for all papers, no matter what type or initiation. In established and unspecified online communities, the outcome was that patients experienced (1) increased participation during health care consultation; (2) awareness of roles, such as when and how to contact and gain better outcomes from consultation; and (3) becoming more informed and having up-to-date knowledge about treatments, care process, and control of emotional management of the condition, which indicated an increased level of self-care (Table 4).

Table 4. The outcome of patient involvement and participation in established and undefined communities.

Characteristic	Established online community		Undefined online community	
	Papers, n (%)[a]	References	Papers, n (%)[a]	References
Total papers (n=28)	21 (64)	[37-57]	7 (21)	[58-64]
Outcome				
Increased participation during consultations	10 (30)	[37,39,40,44-47,51,54,57]	4 (12)	[58-60,64]
Awareness of care trajectory and how to gain better outcomes from consultation	10 (30)	[37,39,40,44-47,51,54,57]	4 (12)	[58-60,64]
Increased level of self-care	20 (61)	[37-52,54-57]	7 (21)	[58-64]
Response by health care professionals				
Positive	3 (9)	[45,48,49]	2 (6)	[58,59]
Negative	7 (21)	[43-45,48,49,55,57]	2 (6)	[58,59]

[a]All percentages refer to the total number of papers, N=33.

Response by Health Care

In 9 out of 33 established or undefined community papers (27%), patients perceived themselves as being more up-to-date via online communities than they felt health care professionals were [40,43,44,46,49-51,58,64]. In another 9 established or undefined community papers (27%), this is described as a positive response by health care professionals, and sometimes, as the opposite [43-45,48,49,55,57-59] (Table 4). If health care professionals had a positive response to usage, patients often experienced themselves as being increasingly involved and having a collaborative relationship with health care professionals. Consequently, this resulted in patients experiencing better navigation in online communities, which affected how to incorporate information that was relevant to their individual situation. If health care professionals had a negative response toward patients' usage of online communities, the consequence was often described as patients' experiencing not being involved during consultation. Instead, the responsibility was all in the hands of the health care professional and was described as being not satisfying for patients (7/33, 21%) [43,44,48,49,55,58,59]. In 4 papers (12%), this was described as a reason for asking patient peers instead of health care professionals for consultation [43,44,55,57], or decreasing contact with health care professionals (3/33, 9%) [43,44,57]. If patients did not experience involvement during the first consultation, it affected whether patients chose to share their experiences of using online communities with health care professionals (5/33, 15%) [45,48,49,58,59].

Designed Online Community Papers

In the designed community papers, the outcome from patient involvement or participation was oriented toward leadership. At first as the health care professionals who determined conditions for treatment (3/33, 9%) [65,67,68], while in later steps when patients had more experience using the online community, the outcomes were that patients determined the conditions—how much patients decided to participate during consultation [67]. In one paper [67], the later steps are described as both positive and negative. The positive outcomes were that patients had more understanding of individual responsibility for health conditions, thus were more self-reliant on management and became more involved during health care consultations. The negative outcomes were that many health care professionals experienced pressure to be available online 24/7, in order to respond to patient contact inquiries. The contact inquiries were mostly regarding turmoil that had emerged during patients' use of the online community. There was a mismatch between the patients' needs and the time the health care professionals had for this type of work, and the health care professionals experienced that some patients did not consider professionals' life outside work or understand that they had other patients to care for, and therefore, had limited time.

Discussion

Summary

This systematic review's objective research question was "In what ways can participation in online communities support patient empowerment?" The findings indicated that participation in online communities, regardless of type, can be seen as a complementary resource to traditional health care, since communities helped patients get more out of the consultation with health care professionals by understanding when to contact or getting an insight from peers (into the whole care trajectory and what to expect at different phases). Therefore, online communities supported patient empowerment by helping the patients become engaged and have the possibility of being equal contributors in the patient-provider relationship [5,6,10]. Additionally, participation in online communities supported patients in healing the emotional wounds of a diagnosis or handling negative experiences of their care trajectory. The emotional and personal experiences seemed to be an essential factor behind patients becoming empowered, thus an online community was a space for dealing with these types of experiences. These findings are relevant since they indicate that the progression to self-care must include personal elements and spaces for dealing with diagnosis [73-75]. This seems to be a limited service given by health care, according to the patient needs identified in the papers that were included.

Even if relevant, the way online communities specifically support patient empowerment is complex and dependent on patients' levels of health literacy and previous online community

experiences. Therefore, we (1) discuss different types of empowerment identified; (2) present limitations with the papers that were included, and simultaneously, give suggestions for future research; (3) propose a framework that can be used for understanding or evaluating in which way participation in online communities could support patients empowerment levels and potential progression, and (4) present limitations in conducting this systematic review, how it might have affected the inclusion of papers and the findings, and recommendations for future research concerning how to improve future conduct of systematic reviews.

Different Types of Empowerment

One of this systematic review's contributions is identifying how participation in online communities supported patient empowerment as both a process and an outcome, which echoes results from and ideas in previous research [3,70,72,76]. The processes and outcomes that are supported depend on initiative and motivation to use online communities; hence there is an importance in unpacking the underlying factors for the way online communities supported patient empowerment. This is similar to identifications made by previous research regarding defining or evaluating patient empowerment in a traditional care trajectory [5,6,9], but also through online communities and other eHealth technologies [3,70,72,76,77].

Patient empowerment processes are often defined as continuously taking part in various forms of empowerment [6]. The systematic review confirms this and shows that these processes are identified through patient enablement, activation and engagement, via support from patient peers, and do not explicitly involve health care professionals. These processes include becoming better informed, receiving and giving emotional support by sharing relatable experiences of living with the diagnosis, helping others, and networking—which are best enhanced by peers. The outcomes identified through patient activation or involvement and participation were also considered suitable to be supported by patient peers rather than by health care professionals. The outcomes of becoming more active included patients' experiences of being better informed, thus affected taking an increasingly leading role during health consultations and was seen through independence shown in self-care, adherence to treatment, acceptance of the diagnostic situation, feelings of control, emotional health, and self-efficacy. These examples are in line with those in previous research [3,70,72,76,77]; however, our paper adds an in-depth understanding of the differences in what these different concepts entail and how they are interrelated. In some papers that were included, the overall outcomes were discussed as leading to collective empowerment, where the aim was to gain collective knowledge within a community in order to make changes to health care services, systems, or ways of financing health care. Therefore, this presents opportunities for online communities to support patient progress to specific levels of empowerment [78]. Collective empowerment through online communities has been a recurring topic of interest in previous research regarding empowerment of employee, individual or consumer motivation to create change within an organization, community or business [79-81] but is not explored as often in online communities for health care purposes [78].

Limitations in the Selected Papers and Future Research

Suggestions for future work is through 3 main areas of interest that concerns the identified limitations of the papers that were included: (1) study recruitment methods, (2) involvement of health care perspective, and (3) measurement of patient empowerment.

Study Recruitment Methods

Most papers that studied established online communities recruited respondents via sharing links in online communities, such as a questionnaire or interview request [38,39,41-43,45-47,50-52,56]. Consequently, this resulted in mostly positive respondents since they already used the online community enough to see the link. There were thereby both issues of sampling and of demographic characteristics (ie, well-educated, with good experience of using the internet, and high health literacy) that resulted; however, papers discussed these as limitations. Therefore, more diverse recruitment methods and a wider selection of respondents is crucial for future research.

Skills Required When Using Online Communities

Despite the limitations regarding recruitment, the findings could be perceived as patients' needing previous experience and skills using online communities in order for the patient to be empowered and fulfill the purpose behind usage. There were patients that had negative experiences of using online communities if their needs were not met as expected. Another important factor was if the amount of information became too much to handle and was based on the individual phase of the diagnostic journey. Usually, the success of online communities depends on members' previous experiences with using internet-related services, the functionality of technology, and their motivation for becoming a member in the first place [21-23]. Therefore, it is important for future research to consider a variety of respondents in order to understand how knowledge and design in online communities could be adapted and evaluated to those who have less prior experience in order for online communities to be beneficial and in order to reduce the risk of digital divide [3,18].

Involvement of a Health Care Perspective

There is a gap in involving a health care perspective. By this we refer to involvement such as integrating health care professionals' views in order to understand how patients have been empowered when participating in online communities and to understand how online communities can be used and acknowledged as a complement to traditional health care. In all papers that were included, there were recommendations for professionals to be better involved and positive toward patient usage. However, only one paper [48] explicitly described how this could be managed. In order for professionals to see the potential of online communities or follow recommendations from research, there need to be strategies on how the use of established and designed online communities should be practiced in daily work and service routine [74,82].

Another limitation was difficulty in identifying level of participation of health care professionals contributing to content

in online communities and how it affected patients. Patients and professionals are considered complementary actors of the patient empowerment experience [5,6]; therefore, the health care professionals' perspective in relation to patient usage should be more involved and highlighted in online community research. This limitation identified in the papers that were included might depend on the conduct of search strategy and inclusion and exclusion criteria of this systematic review (Table S1, Multimedia Appendix 2).

Measurement of Patient Empowerment

Measurement and definition of patient empowerment were also a limitation in selected papers. In inductive approach papers, patient empowerment was mainly measured through respondents stating they felt empowerment, while in deductive papers through measurements. There were many individual adaptations of measurement tools, which made it difficult to evaluate the quality of studies [19]. These limitations depend on patient empowerment as a concept that has many different definitions and interpretations depending on context and diagnosis [5,6,10,67,72-74]. Another reason can be that there are few validated analytical frameworks that can be used when evaluating health-related effects in online health communities [19]. Therefore, we propose a revised framework that aims to evaluate different levels of patient empowerment, and progression made through online communities. This framework was developed during the analysis process.

Framework of Patient Empowerment Levels

Based on the findings in this systematic review in relation to the limitations identified, we propose a revised framework for empowerment consisting of hierarchical levels (Table 5). These hierarchical levels can be used to identify in which way an online community can support patient empowerment processes or to construct trajectories of progress based on the patients' needs and where the patients are on their journey. Additionally, the framework can give guidance in how to methodically identify hierarchical levels of patient empowerment. By hierarchical levels we mean levels that can describe phases that patients are currently in, which can make it easier to evaluate paths for progression. There is a general consensus regarding progression as part of the empowerment concept, regardless of which concepts of patient empowerment are being used to discuss it [1,6,7,10,79-81]. Therefore, when discussing online communities in relation to patient empowerment, progression within specific phases is an important aspect. The progression affects the way the patients perceive their own empowerment and how susceptible they are to motivation and effects of self-care, their feelings of control, and the way they conduct contact with health care professionals. It is also important to keep in mind that not all patients want or have the ability to progress.

Additionally, the level of empowerment also affects how patients use online communities as support in the process of progression and how they support others, which in turn affects the collective empowerment (Table 5). Identifying where a patient is in the progression process is therefore an important aspect and the framework can be used to do so. However, it is important to keep in mind that people who have satisfied their informational and support needs may choose to leave the community or do not want to continue their empowerment progression by becoming producers or patient mentors in the online community. The reason might be that they believe to have completed their diagnostic journey, thus do not want to be reminded of their previous situation [83]. Another aspect to consider before using this type of framework is the nonlinearity of potential empowerment progression, that can happen if, for example, a new diagnosis or something else happens that affect management of condition and trajectory. Another example to consider is that several levels could occur simultaneously and might depend on the structure of the online community [21-23]. No matter what, the most important aspect to keep in mind is that not all patients want or have the ability to progress; therefore, we recommend that others evaluate this framework with caution.

Table 5. The proposed hierarchical framework of patient empowerment levels.

Hierarchical level of patient empowerment	Definition of level
Level 1: Motivated patient	The motivated patient is motivated to adhere to treatment and information given by the professional but lets the professional take the leading role within health care consultation.
Level 2: Self-cared patient	The self-cared patient takes control over disease and seeks information and knowledge that will help improve self-management or takes a leading role in health consultation. The patient uses the online community in order to get a second opinion or support potential void of information or knowledge from health care, regarding emotional or social aspects of living with the disease. At this level, the patient is thus more driven and uses online communities as complements as to traditional health care and has taken ownership of the disease.
Level 3: Producing patient	The producing patient means that the patient is not just a passive consumer of care since the patient wants to help others by sharing their experiences with disease and health care process. The context behind helping others might depend on the patient wanting to learn, and simultaneously gain status and satisfaction, in order to improve emotional or mental health. Additionally, the context of learning and gaining status or satisfaction and how it improves emotional or mental health could be an outcome of helping others.
Level 4: Patient activist	Collective empowerment is considered to be the patient having the aim of informing or helping others in order to develop or change policies and awareness in health care. Patients' experiences become evidence-based and used by health care professionals.

Limitations of This Work

Way of Reporting Included Papers

The main characteristics of papers that were included were individually reported (Multimedia Appendix 6). However, full descriptions of details such as country of origin, the sample size of respondents, and the online community's domain were not always presented. Papers differed with their transparency, which made it difficult to report consistently and took a lot of time. Hence, we decided to exclude categories typically reported in systematic reviews, which may limit the transparency and quality of our findings. Therefore, we recommend future research to divide the work of systematic review into phases of looking into specific details of the papers that were included and reporting them in a particular time period in order to make it time-efficient and simultaneously maintain the quality of work that is expected of systematic reviews [24,34,84].

No Quality Rating

It could be argued that it is standard to use quality rating of papers that were included in systematic reviews to evaluate risk of bias and how to evaluate the validity of the findings identified [24,84]. The reason for not including quality rating in this systematic review was based on wanting to focus on how to visualize the analysis of patient empowerment in order to make potential contribution of making the concept clear when put into specific context [5,6,9]. Additionally, it was difficult to evaluate which type of quality rating standard should be used since the papers that were included varied (in measuring empowerment and context of studies; Multimedia Appendix 6). This difficulty depended on the objective, time limitation, and not following a linear process in planning and conducting the systematic review. Therefore, we recommend future research to carefully structure the planning phase of doing systematic reviews in order to follow the principles of the PRISMA statement and rules of conduct when doing high-quality systematic reviews.

Conclusion and Future of Patient Empowerment in Online Communities

This systematic review shows in which ways participation in online communities could support patient empowerment. The main findings indicated that online communities supported patient empowerment in the way of meeting emotional need of handling condition and the possibility of patients becoming equal contributors to the patient-provider relationship. An additional finding was that online communities supported both process and outcomes of patient empowerment. The main contribution of this systematic review is a framework and conceptualization of how patient empowerment in online communities can be understood, evaluated, and designed for empowerment progression and support. Based on identification of main findings, we suggest that future work look specifically toward 3 main areas of interest: (1) study recruitment methods; (2) involvement of a health care perspective; and (3) measurement of patient empowerment. Based on all suggestions, we propose that our framework can be used to evaluate different levels of patient empowerment and progression through online communities.

Acknowledgments

Funding was provided by the Swedish Cancer Society (Senior Clinical Investigator Award; CAN 2016/509), the Swedish Research Council (2017-01103), the agreement concerning research and education of doctors (ALFGBG-716581) and the Region Västra Götaland Healthcare Committee (*Hälso- och sjukvårdsstyrelsen*; VGFOUREG-855181 and 931677).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Report of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.
[DOC File , 65 KB - jmir_v23i2e19910_app1.doc]

Multimedia Appendix 2
Inclusion and exclusion criteria.
[DOCX File , 23 KB - jmir_v23i2e19910_app2.docx]

Multimedia Appendix 3
Database searches.
[DOCX File , 24 KB - jmir_v23i2e19910_app3.docx]

Multimedia Appendix 4
Analytical framework of patient empowerment.
[DOCX File , 17 KB - jmir_v23i2e19910_app4.docx]

Multimedia Appendix 5

Excluded papers.
[XLSX File (Microsoft Excel File), 204 KB - jmir_v23i2e19910_app5.xlsx]

Multimedia Appendix 6
General characteristics of selected papers.
[XLSX File (Microsoft Excel File), 19 KB - jmir_v23i2e19910_app6.xlsx]

References

1. Rappaport J. Empowerment meets narrative: listening to stories and creating settings. Am J Community Psychol 1995 Oct 01;23(5):795-807 [FREE Full text] [doi: 10.1007/bf02506992]
2. Yan L, Peng J, Tan Y. Network dynamics: how can we find patients like us? Inf Syst Res 2015 Sep 11;26(3):496-512 [FREE Full text] [doi: 10.1287/isre.2015.0585]
3. Smailhodzic E, Hooijsma W, Boonstra A, Langley DJ. Social media use in healthcare: a systematic review of effects on patients and on their relationship with healthcare professionals. BMC Health Serv Res 2016 Aug 26;16:442 [FREE Full text] [doi: 10.1186/s12913-016-1691-0] [Medline: 27562728]
4. Eysenbach G. The impact of the internet on cancer outcomes. CA Cancer J Clin 2003;53(6):356-371 [FREE Full text] [doi: 10.3322/canjclin.53.6.356] [Medline: 15224975]
5. Fumagalli LP, Radaelli G, Lettieri E, Bertele' P, Masella C. Patient Empowerment and its neighbours: clarifying the boundaries and their mutual relationships. Health Policy 2015 Mar;119(3):384-394. [doi: 10.1016/j.healthpol.2014.10.017] [Medline: 25467286]
6. Palumbo R. The Bright Side and the Dark Side of Patient Empowerment: Co-creation and co-destruction of Value in the Healthcare Environment. Cham, Switzerland: Springer International Publishing; May 31, 2017.
7. Zimmerman M. Psychological empowerment: issues and illustrations. Am J Community Psychol 1995 Oct 01;23(5):581-599 [FREE Full text] [doi: 10.1007/BF02506983] [Medline: 8851341]
8. Ekman I, Swedberg K, Taft C, Lindseth A, Norberg A, Brink E, et al. Person-centered care--ready for prime time. Eur J Cardiovasc Nurs 2011 Dec 01;10(4):248-251 [FREE Full text] [doi: 10.1016/j.ejcnurse.2011.06.008] [Medline: 21764386]
9. Holmström I, Röing M. The relation between patient-centeredness and patient empowerment: a discussion on concepts. Patient Educ Couns 2010 May;79(2):167-172 [FREE Full text] [doi: 10.1016/j.pec.2009.08.008] [Medline: 19748203]
10. Anderson RM, Funnell MM. Patient empowerment: myths and misconceptions. Patient Educ Couns 2010 Jun;79(3):277-282 [FREE Full text] [doi: 10.1016/j.pec.2009.07.025] [Medline: 19682830]
11. Hansson E, Ekman I, Swedberg K, Wolf A, Dudas K, Ehlers L, et al. Person-centred care for patients with chronic heart failure - a cost-utility analysis. Eur J Cardiovasc Nurs 2016 Jun 01;15(4):276-284 [FREE Full text] [doi: 10.1177/1474515114567035] [Medline: 25595358]
12. Olsson L, Hansson E, Ekman I, Karlsson J. A cost-effectiveness study of a patient-centred integrated care pathway. J Adv Nurs 2009 Aug;65(8):1626-1635. [doi: 10.1111/j.1365-2648.2009.05017.x] [Medline: 19493145]
13. Pirhonen L, Bolin K, Olofsson EH, Fors A, Ekman I, Swedberg K, et al. Person-centred care in patients with acute coronary syndrome: cost-effectiveness analysis alongside a randomised controlled trial. Pharmacoecon Open 2019 Dec;3(4):495-504 [FREE Full text] [doi: 10.1007/s41669-019-0126-3] [Medline: 30825129]
14. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health Aff (Millwood) 2009;28(2):361-368. [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
15. Li C, Bernhoff J. Groundswell: Winning In A World Transformed By Social Media Technologies. Boston: Harvard Business Press; May 01, 2011.
16. Williams J. A new road map for healthcare business success. Healthc Financ Manage 2011 May;65(5):62-69. [Medline: 21634269]
17. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
18. Maclachlan LR, Mills K, Lawford BJ, Egerton T, Setchell J, Hall LM, et al. Design, delivery, maintenance, and outcomes of peer-to-peer online support groups for people with chronic musculoskeletal disorders: systematic review. J Med Internet Res 2020 Apr 24;22(4):e15822 [FREE Full text] [doi: 10.2196/15822] [Medline: 32329746]
19. Griffiths KM, Calear AL, Banfield M. Systematic review on internet support groups (ISGs) and depression (1): do ISGs reduce depressive symptoms? J Med Internet Res 2009 Sep 30;11(3):e40 [FREE Full text] [doi: 10.2196/jmir.1270] [Medline: 19793719]
20. van Eenbergen MC, van de Poll-Franse LV, Heine P, Mols F. The impact of participation in online cancer communities on patient reported outcomes: systematic review. JMIR Cancer 2017 Sep 28;3(2):e15 [FREE Full text] [doi: 10.2196/cancer.7312] [Medline: 28958985]
21. Preece J. Sociability and usability in online communities: determining and measuring success. Behav Inf Technol 2010 Nov 08;20(5):347-356 [FREE Full text] [doi: 10.1080/01449290110084683]

22. Skousen T, Safadi H, Young C, Karahanna E, Safadi S, Chebib F. Successful moderation in online patient communities: inductive case study. J Med Internet Res 2020 Mar 17;22(3):e15983 [FREE Full text] [doi: 10.2196/15983] [Medline: 32181743]
23. Young C. Community management that works: how to build and sustain a thriving online health community. J Med Internet Res 2013 Jun 11;15(6):e119 [FREE Full text] [doi: 10.2196/jmir.2501] [Medline: 23759312]
24. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
25. Enam A, Torres-Bonilla J, Eriksson H. Evidence-based evaluation of eHealth interventions: systematic literature review. J Med Internet Res 2018 Nov 23;20(11):e10971 [FREE Full text] [doi: 10.2196/10971] [Medline: 30470678]
26. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
27. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117 [FREE Full text] [doi: 10.2196/13117] [Medline: 31115340]
28. Hong YA, Liang C, Radcliff TA, Wigfall LT, Street RL. What do patients say about doctors online? a systematic review of studies on patient online reviews. J Med Internet Res 2019 Apr 08;21(4):e12521 [FREE Full text] [doi: 10.2196/12521] [Medline: 30958276]
29. Spencer L, Potterton R, Allen K, Musiat P, Schmidt U. Internet-based interventions for carers of individuals with psychiatric disorders, neurological disorders, or brain injuries: systematic review. J Med Internet Res 2019 Jul 09;21(7):e10876. [doi: 10.2196/10876] [Medline: 31290399]
30. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
31. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
32. Preece J, Maloney-Krichmar D, Abras C. History of online communities. In: Christensen K, Levin D, editors. Encyclopedia of Community: From Village to Virtual World. Thousand Oaks: Sage Publications; 2003:1023-1027.
33. Bryman A. Social Research Methods. Oxford: Oxford University Press; Feb 01, 2016.
34. Webster J, Watson R. Analyzing the past to prepare for the future: writing a literature review. MIS quarterly 2002 Jun;26(2):xiii-xxiii [FREE Full text]
35. Knopf JW. Doing a literature review. PS: Political Science and Politics 2006;39(01):127-132 [FREE Full text]
36. Synthesis matrix. The University of Arizona. 2019. URL: https://writingcenter.ashford.edu/synthesis-matrix [accessed 2020-01-02]
37. Aardoom JJ, Dingemans AE, Boogaard LH, Van Furth EF. Internet and patient empowerment in individuals with symptoms of an eating disorder: a cross-sectional investigation of a pro-recovery focused e-community. Eat Behav 2014 Aug;15(3):350-356 [FREE Full text] [doi: 10.1016/j.eatbeh.2014.04.003] [Medline: 25064280]
38. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Self-reported differences in empowerment between lurkers and posters in online patient support groups. J Med Internet Res 2008 Jun 30;10(2):e18 [FREE Full text] [doi: 10.2196/jmir.992] [Medline: 18653442]
39. van Uden-Kraan C, Drossaert C, Taal E, Seydel E, van de Laar M. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009 Jan;74(1):61-69 [FREE Full text] [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
40. Ammari T, Schoenebeck S. Networked empowerment on Facebook groups for parents of children with special needs. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. New York, USA: Association for Computing Machinery; 2015 Apr Presented at: CHI: Conference on Human Factors in Computing Systems; 2015; Seoul Republic of Korea p. 2805-2814 URL: https://dl.acm.org/doi/abs/10.1145/2702123.2702324 [doi: 10.1145/2702123.2702324]
41. Batenburg A, Das E. Emotional approach coping and the effects of online peer-led support group participation among patients with breast cancer: a longitudinal study. J Med Internet Res 2014 Nov 28;16(11):e256 [FREE Full text] [doi: 10.2196/jmir.3517] [Medline: 25474819]
42. Batenburg A, Das E. Virtual support communities and psychological well-being: the role of optimistic and pessimistic social comparison strategies. J Comput-Mediat Comm 2015 Aug 24;20(6):585-600 [FREE Full text] [doi: 10.1111/jcc4.12131]
43. Coulson NS, Shaw RL. Nurturing health-related online support groups: exploring the experiences of patient moderators. Comput Human Behav 2013 Jul;29(4):1695-1701 [FREE Full text] [doi: 10.1016/j.chb.2013.02.003]
44. Fernandes LDS, Calado C, Araujo CAS. Social networks and health practices: influence of a diabetes online community on adherence to treatment. Cien Saude Colet 2018 Oct;23(10):3357-3368 [FREE Full text] [doi: 10.1590/1413-812320182310.14122018] [Medline: 30365855]

45. Grosberg D, Grinvald H, Reuveni H, Magnezi R. Frequent Frequent surfing on social health networks is associated with increased knowledge and patient health activation. J Med Internet Res 2016 Aug 10;18(8):e212 [FREE Full text] [doi: 10.2196/jmir.5832] [Medline: 27511272]
46. Harkin LJ, Beaver K, Dey P, Choong K. Navigating cancer using online communities: a grounded theory of survivor and family experiences. J Cancer Surviv 2017 Dec;11(6):658-669 [FREE Full text] [doi: 10.1007/s11764-017-0616-1] [Medline: 28470506]
47. Johnston AC, Worrell JL, Di Gangi PM, Wasko M. Online health communities: an assessment of the influence of participation on patient empowerment outcomes. Info Technology & People 2013 May 31;26(2):213-235 [FREE Full text] [doi: 10.1108/itp-02-2013-0040]
48. Litchman ML, Edelman LS, Donaldson GW. Effect of diabetes online community engagement on health indicators: cross-sectional study. JMIR Diabetes 2018 Apr 24;3(2):e8 [FREE Full text] [doi: 10.2196/diabetes.8603] [Medline: 30291079]
49. Litchman ML, Rothwell E, Edelman LS. The diabetes online community: Older adults supporting self-care through peer health. Patient Educ Couns 2018 Mar;101(3):518-523 [FREE Full text] [doi: 10.1016/j.pec.2017.08.023] [Medline: 28947360]
50. Mo PKH, Coulson NS. Developing a model for online support group use, empowering processes and psychosocial outcomes for individuals living with HIV/AIDS. Psychol Health 2012;27(4):445-459 [FREE Full text] [doi: 10.1080/08870446.2011.592981] [Medline: 21854088]
51. Mo PKH, Coulson NS. Are online support groups always beneficial? a qualitative exploration of the empowering and disempowering processes of participation within HIV/AIDS-related online support groups. Int J Nurs Stud 2014 Jul;51(7):983-993. [doi: 10.1016/j.ijnurstu.2013.11.006] [Medline: 24342449]
52. Oh HJ, Lee B. The effect of computer-mediated social support in online communities on patient empowerment and doctor-patient communication. Health Commun 2012;27(1):30-41. [doi: 10.1080/10410236.2011.567449] [Medline: 21797714]
53. Petrovčič A, Petrič G. Differences in intrapersonal and interactional empowerment between lurkers and posters in health-related online support communities. Comput Human Behav 2014 May;34:39-48 [FREE Full text] [doi: 10.1016/j.chb.2014.01.008]
54. Pousti H, Urquhart C, Linger H. Exploring the role of social media in chronic care management: a socialmaterial approach. In: Doolin B, Lamprou E, Mitev N, McLeod L, editors. IFIP Advances in Information and Communication Technology. Berlin, Heidelberg: Springer; 2014:163-185.
55. Rubenstein EL. Rituals of introduction and revolving roles: socialization in an online breast cancer community. Libr Inf Sci Res 2015 Oct;37(4):353-362 [FREE Full text] [doi: 10.1016/j.lisr.2015.11.005]
56. Wicks P, Keininger DL, Massagli MP, de la Loge C, Brownstein C, Isojärvi J, et al. Perceived benefits of sharing health data between people with epilepsy on an online platform. Epilepsy Behav 2012 Jan;23(1):16-23 [FREE Full text] [doi: 10.1016/j.yebeh.2011.09.026] [Medline: 22099528]
57. Zhang R, Eschler J, Reddy M. Online support groups for depression in China: culturally shaped interactions and motivations. Comput Supported Coop Work 2018 May 15;27(3-6):327-354 [FREE Full text] [doi: 10.1007/s10606-018-9322-4]
58. Bartlett YK, Coulson NS. An investigation into the empowerment effects of using online support groups and how this affects health professional/patient communication. Patient Educ Couns 2011 Apr;83(1):113-119 [FREE Full text] [doi: 10.1016/j.pec.2010.05.029] [Medline: 20599338]
59. Brady E, Segar J, Sanders C. Accessing support and empowerment online: the experiences of individuals with diabetes. Health Expect 2017 Oct;20(5):1088-1095 [FREE Full text] [doi: 10.1111/hex.12552] [Medline: 28718928]
60. Reeves PM. Coping in cyberspace: the impact of internet use on the ability of HIV-positive individuals to deal with their illness. J Health Commun 2000;5 Suppl:47-59. [doi: 10.1080/10810730050019555] [Medline: 11010356]
61. Santana S, Lausen B, Bujnowska-Fedak M, Chronaki CE, Prokosch H, Wynn R. Informed citizen and empowered citizen in health: results from an European survey. BMC Fam Pract 2011 Apr 16;12(1):20 [FREE Full text] [doi: 10.1186/1471-2296-12-20] [Medline: 21496309]
62. Seçkin G. Informational and decisional empowerment in online health support communities: initial psychometric validation of the Cyber Info-Decisional Empowerment Scale (CIDES) and preliminary data from administration of the scale. Support Care Cancer 2011 Dec;19(12):2057-2061. [doi: 10.1007/s00520-011-1249-y] [Medline: 21858718]
63. Seçkin G. Using hierarchical multivariate analysis to examine interactions between alternative appraisals of cancer and virtual health support among middle-aged and older internet users. Res Aging 2012 Oct 29;35(6):688-709. [doi: 10.1177/0164027512460692]
64. Wentzer HS, Bygholm A. Narratives of empowerment and compliance: studies of communication in online patient support groups. Int J Med Inform 2013 Dec;82(12):e386-e394. [doi: 10.1016/j.ijmedinf.2013.01.008] [Medline: 23523082]
65. Allam A, Kostova Z, Nakamoto K, Schulz PJ. The effect of social support features and gamification on a web-based intervention for rheumatoid arthritis patients: randomized controlled trial. J Med Internet Res 2015 Jan 09;17(1):e14 [FREE Full text] [doi: 10.2196/jmir.3510] [Medline: 25574939]

66. Brighton LJ, Pask S, Benalia H, Bailey S, Sumerfield M, Witt J, et al. Taking patient and public involvement online: qualitative evaluation of an online forum for palliative care and rehabilitation research. Res Involv Engagem 2018 May 1;4(1):14 [FREE Full text] [doi: 10.1186/s40900-018-0097-z] [Medline: 29744131]
67. Gammon D, Strand M, Eng LS, Børøsund E, Varsi C, Ruland C. Shifting practices toward recovery-oriented care through an e-recovery portal in community mental health care: a mixed-methods exploratory study. J Med Internet Res 2017 May 02;19(5):e145 [FREE Full text] [doi: 10.2196/jmir.7524] [Medline: 28465277]
68. Geramita EM, Herbeck Belnap B, Abebe KZ, Rothenberger SD, Rotondi AJ, Rollman BL. The association between increased levels of patient engagement with an internet support group and improved mental health outcomes at 6-month follow-up: post-hoc analyses from a randomized controlled trial. J Med Internet Res 2018 Jul 17;20(7):e10402 [FREE Full text] [doi: 10.2196/10402] [Medline: 30021711]
69. Kaplan K, Salzer MS, Solomon P, Brusilovskiy E, Cousounis P. Internet peer support for individuals with psychiatric disabilities: A randomized controlled trial. Soc Sci Med 2011 Jan;72(1):54-62. [doi: 10.1016/j.socscimed.2010.09.037] [Medline: 21112682]
70. van Uden-Kraan CF, Drossaert CHC, Taal E, Shaw BR, Seydel ER, van de Laar MAFJ. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. Qual Health Res 2008 Mar;18(3):405-417 [FREE Full text] [doi: 10.1177/1049732307313429] [Medline: 18235163]
71. Spreitzer GM. Psychological empowerment in the workplace: dimensions, measurement, and validation. AMJ 1995 Oct;38(5):1442-1465 [FREE Full text] [doi: 10.5465/256865]
72. Barak A, Boniel-Nissim M, Suler J. Fostering empowerment in online support groups. Comput Human Behav 2008 Sep;24(5):1867-1883 [FREE Full text] [doi: 10.1016/j.chb.2008.02.004]
73. Dentzer S. Health Care Without Walls: A Roadmap for Reinventing US Health Care. Boston, USA: Network for Excellence in Health Innovation; 2019:A.
74. Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. Per Med 2013;10(6):565-576 [FREE Full text] [doi: 10.2217/pme.13.57] [Medline: 25342952]
75. Hesse BW. Role of the internet in solving the last mile problem in medicine. J Med Internet Res 2019 Oct 28;21(10):e16385 [FREE Full text] [doi: 10.2196/16385] [Medline: 31661078]
76. Kingod N, Cleal B, Wahlberg A, Husted GR. Online peer-to-peer communities in the daily lives of people with chronic illness: a qualitative systematic review. Qual Health Res 2017 Jan;27(1):89-99. [doi: 10.1177/1049732316680203] [Medline: 27956659]
77. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating patient empowerment in association with eHealth technology: scoping review. J Med Internet Res 2017 Sep 29;19(9):e329 [FREE Full text] [doi: 10.2196/jmir.7809] [Medline: 28963090]
78. Atanasova S, Petric G. Collective empowerment in online health communities: scale development and empirical validation. J Med Internet Res 2019 Nov 20;21(11):e14392 [FREE Full text] [doi: 10.2196/14392] [Medline: 31746772]
79. Drury J, Reicher S. The intergroup dynamics of collective empowerment: Substantiating the social identity model of crowd behavior. Group Process Intergroup Relat 2016 Jul 25;2(4):381-402 [FREE Full text] [doi: 10.1177/1368430299024005]
80. Joo D, Woosnam KM, Strzelecka M, Boley BB. Knowledge, empowerment, and action: testing the empowerment theory in a tourism context. Journal of Sustainable Tourism 2019 Oct 29;28(1):69-85 [FREE Full text] [doi: 10.1080/09669582.2019.1675673]
81. Sjöberg S, Rambaree K, Jojo B. Collective empowerment: a comparative study of community work in Mumbai and Stockholm. Int J Soc Welf 2014 Dec 09;24(4):364-375 [FREE Full text] [doi: 10.1111/ijsw.12137]
82. Vallo Hult H, Hansson A, Svensson L, Gellerstedt M. Flipped healthcare for better or worse. Health Informatics J 2019 Sep;25(3):587-597 [FREE Full text] [doi: 10.1177/1460458219833099] [Medline: 30887867]
83. Islind A, Johansson V, Vallo Hult H, Alsén P, Andreasson E, Angenete E, et al. Individualized blended care for patients with colorectal cancer: the patient's view on informational support. Support Care Cancer 2020 Oct 12:1-7 [FREE Full text] [doi: 10.1007/s00520-020-05810-5] [Medline: 33044626]
84. Waddington H, White H, Snilstveit B, Hombrados JG, Vojtkova M, Davies P, et al. How to do a good systematic review of effects in international development: a tool kit. J Dev Effect 2012 Sep;4(3):359-387 [FREE Full text] [doi: 10.1080/19439342.2012.711765]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 06.05.20; peer-reviewed by R Palumbo, C Young, B Rollman; comments to author 26.10.20; revised version received 14.12.20; accepted 09.01.21; published 09.02.21.

Please cite as:
Johansson V, Islind AS, Lindroth T, Angenete E, Gellerstedt M
Online Communities as a Driver for Patient Empowerment: Systematic Review
J Med Internet Res 2021;23(2):e19910
URL: http://www.jmir.org/2021/2/e19910/
doi:10.2196/19910
PMID:33560233

Review

Effects of eHealth-Based Multiple Health Behavior Change Interventions on Physical Activity, Healthy Diet, and Weight in People With Noncommunicable Diseases: Systematic Review and Meta-analysis

Yanping Duan[1], PhD; Borui Shang[2], PhD; Wei Liang[1], PhD; Gaohui Du[3], MEd; Min Yang[1], MEd; Ryan E Rhodes[4], PhD

[1]Department of Sport, Physical Education and Health, Hong Kong Baptist University, Hong Kong, China

[2]Department of Social Sciences, Hebei Sport University, Shijiazhuang, China

[3]Department of Health Science, Wuhan Sports University, Wuhan, China

[4]School of Exercise Science, Physical and Health Education, University of Victoria, Victoria, BC, Canada

Corresponding Author:
Borui Shang, PhD
Department of Social Sciences
Hebei Sport University
82 Xuefu Road
Shijiazhuang, 050041
China
Phone: 86 15383112089
Email: borui_shang_pe@qq.com

Abstract

Background: Noncommunicable diseases (NCDs) are associated with the burden of premature deaths and huge medical costs globally. There is an increasing number of studies combining a multiple health behavior change (MHBC) intervention paradigm with eHealth approaches to jointly promote weight-related health behaviors among people with NCD; yet, a comprehensive summary of these studies is lacking.

Objective: This review aims to meta-analyze the effectiveness and systematically summarize the characteristics of the relevant intervention studies for improving the outcomes of physical activity, healthy diet, and weight among people with NCD.

Methods: Following PRISMA guidelines, 4 electronic databases (PsycINFO, PubMed, Scopus, SPORTDiscus) were systematically searched to identify eligible articles based on a series of inclusion and exclusion criteria. Article selection, quality assessment, and data extraction were independently performed by 2 authors. The standardized mean difference (SMD) was calculated to evaluate the effectiveness of interventions for 3 intervention outcomes (physical activity, healthy diet, and weight), and subsequent subgroup analyses were performed for gender, age, intervention duration, channel, and theory. Calculations were conducted, and figures were produced in SPSS 22 and Review Manager 5.3.

Results: Of the 664 original hits generated by the systematic searches, 15 eligible studies with moderate to high quality were included. No potential publication bias was detected using statistical analyses. Studies varied in intervention channel, intensity, and content. The meta-analysis revealed that the eHealth MHBC interventions significantly promoted physical activity (SMD 0.85, 95% CI 0.23 to 1.47, *P*=.008) and healthy diet (SMD 0.78, 95% CI 0.13 to 1.43, *P*=.02), but did not contribute to a healthy weight status (SMD –0.13, 95% CI= –0.47 to 0.20, *P*=.43) among people with NCDs, compared to the control conditions. Results from subgroup analysis indicated that theory-based interventions achieved greater effect than nontheory-based interventions in promoting physical activity, and interventions with traditional approaches (SMS, telephone) were more effective than those with modern internet-based approaches in promoting healthy diet.

Conclusions: The results of this review indicates that eHealth MHBC interventions achieve preliminary success in promoting physical activity and healthy diet behaviors among people with NCD. Future studies could improve the intervention design to achieve better intervention effectiveness.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42019118629; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=118629

(J Med Internet Res 2021;23(2):e23786) doi:10.2196/23786

KEYWORDS

systematic review; meta-analysis; noncommunicable disease; multiple health behavior change; weight-related; physical activity; healthy diet; eHealth

Introduction

Noncommunicable diseases (NCDs) pose a major threat to global public health. NCDs, such as cardiovascular diseases, cancers, and diabetes, are the leading causes of death worldwide, causing 41 million deaths each year, equivalent to 71% of all deaths globally [1]. Furthermore, NCD-related medical costs significantly contribute to health care expenditure in many areas around the world [2,3].

For people with NCDs, in addition to obtaining traditional medical treatment, it is essential to adopt a healthy lifestyle through health care intervention in order to avoid further progression and relapse of NCDs [4]. In the recent 15 years, multiple health behavior change (MHBC, namely addressing no less than 2 health behaviors within a limited time period) has demonstrated early success in facilitating a healthy lifestyle among people with NCDs [5-7]. A statement in the Lancet pointed out that weight-related healthy behaviors including regular physical activity (PA) and healthy diet are promising interventions to control the NCD crisis globally [8].

The rationale of applying MHBC to weight-related behavior change is that most of the weight-related unhealthy behaviors (ie, physical inactivity, unhealthy diet) co-occur and are modifiable [6,9,10]. This assumption was empirically supported by a longitudinal study in which an unhealthy diet and physical inactivity strongly contributed to the onset of NCDs and a chain of negative effects including mortality and increasing health care costs [11]. Moreover, it was proposed that “the effect of a small step leading to a big leap forward” also applied to MHBC. A relatively easy health behavior change may serve as a gateway to an overall healthy lifestyle transition, as positive psychological factors such as self-efficacy and motivation can be boosted along with the initial behavior change, which in turn positively affects the subsequent one [9].

In applying MHBC among people with NCDs, a cost-effective mode is to employ up-to-date eHealth approaches by using information and communication technologies [12,13]. With the increasing number of people with NCDs, the traditional face-to-face intervention paradigm can hardly meet the needs of the population with NCDs. Thus eHealth, as an emerging delivery channel for health services and information using the internet and related technologies and media (eg, computers, smartphones), can be a potentially useful supplement for traditional interventions for the population with NCDs in aftercare family settings after discharge [14]. eHealth interventions also break the distance limitation and thus are highly recommended for their low cost, high efficiency, and easy data collection [15]. Many reviews have already shown substantial effects of employing eHealth approaches in addressing a single behavioral domain of either PA or healthy diet among people with NCDs [16,17]. For example, Haberlin et al [16] reviewed the use of eHealth to promote PA in cancer survivors and found that all the 10 included studies reported improvements in PA, with 8 of 10 studies reporting statistically significant changes.

There is an increasing number of studies combining the MHBC intervention paradigm with eHealth approaches to jointly promote weight-related health behaviors (PA and healthy diet) among people with NCDs [18,19]. However, a comprehensive summary of these relevant studies regarding the overall effects and study characteristics is still lacking. To fill this gap, this review mainly aimed to systematically summarize the characteristics of the relevant intervention studies and then pool the effect sizes from the relevant studies to accurately quantify the effects of those interventions on PA, healthy diet, and weight. Findings of this review can provide recommendations for researchers and clinicians to develop effective eHealth intervention programs to promote PA and healthy diet among people with NCDs.

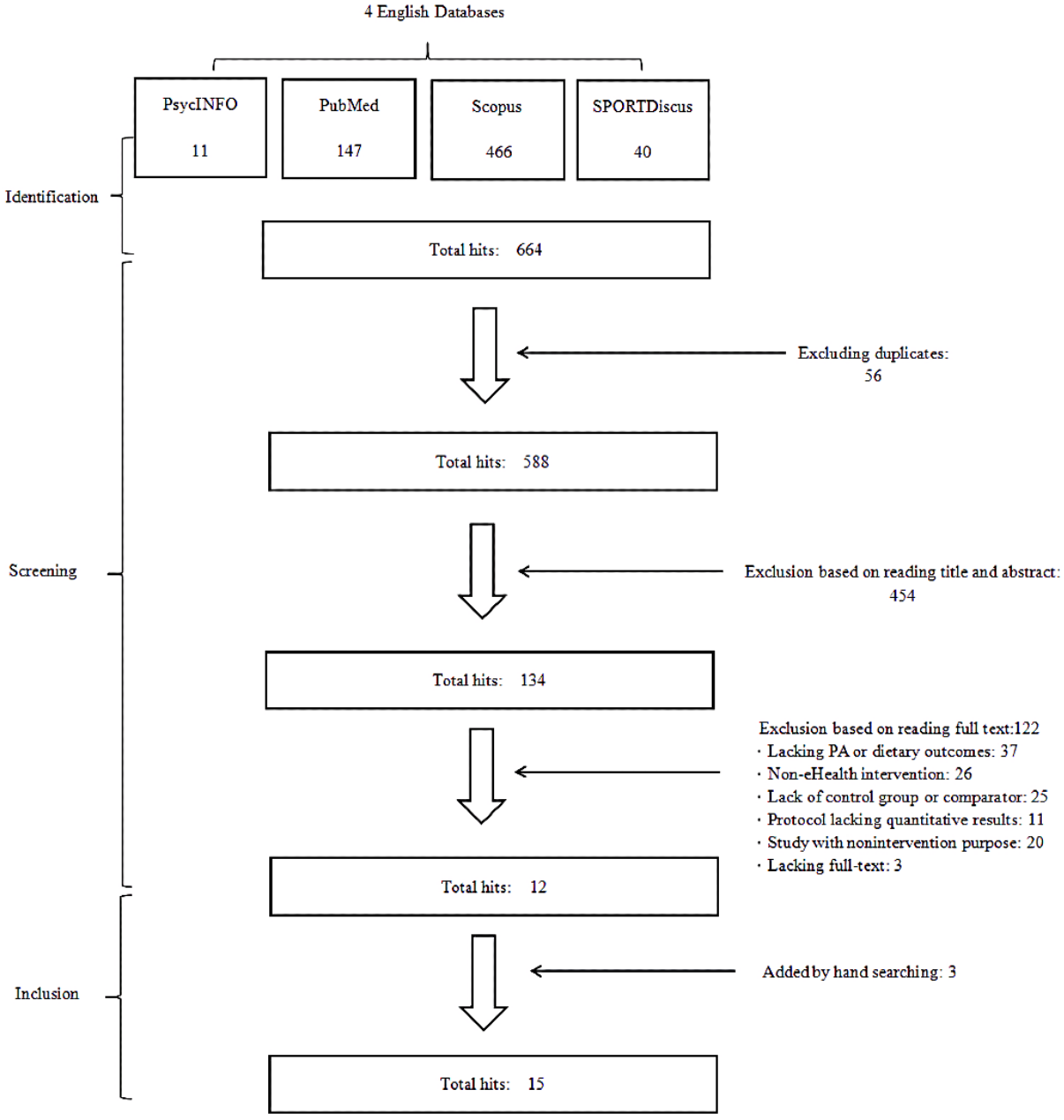
Methods

Search Strategies

This review was conducted and is reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, and the protocol can be retrieved from the PROSPERO database (Registration ID: CRD42019118629) [20]. According to the initial registration, we planned to search relevant articles from both English and Chinese databases. Yet, during the implementation, we found that articles retrieved from Chinese databases were unable to meet the quality standard due to misreporting of critical information (eg, participant characteristics, intervention details) [21]. Thus, we only focused on articles in English databases. A series of structured electronic searches was performed in 4 English databases including PsycINFO, PubMed, Scopus, and SPORTDiscus, focusing on MHBC eHealth interventions regarding weight-related health behaviors (PA and dietary behaviors). The procedures guiding article inclusion are presented in the flow chart in Figure 1. The specific search terms connected with Boolean operators can be seen in Multimedia Appendix 1. The searches were limited to studies with human participants and to publishing dates between 01/01/2000 and 01/03/2020.

All articles identified in the search strategy were exported into reference management software (Mendeley) for duplicate checking and further screening. The reference lists of eligible articles were further reviewed to identify other relevant studies. Relevant reviews that emerged from the search strategy were checked for any additional studies. Grey literature (eg, working papers, unpublished studies, conference proceedings or abstracts, dissertations) was not considered eligible.

Figure 1. PRISMA flow chart of the search strategy and article inclusion.



Study Inclusion and Exclusion Criteria

Type of Participants

The population targeted in this review was people of all genders and any age range with NCDs (eg, cardiovascular diseases, cancers, chronic respiratory diseases, diabetes). Exclusion criteria included studies with participants who were <18 years old, were pregnant or lactating, or had a special condition or other comorbidities that seriously affected their feeding ability and physical mobility (eg, physical disability).

Type of Intervention or Phenomena of Interest

We included studies that evaluated eHealth interventions with the primary aim of affecting behavior change that at least simultaneously incorporated PA and healthy diet behaviors. As such, studies with other irrelevant purposes (eg, investigation of behavioral patterns, investigation of the effectiveness of medical therapy, no eHealth intervention arm) were excluded.

Comparators

Comparators were defined as control groups without an intervention or non-eHealth intervention groups (eg, face-to-face intervention, pamphlet intervention, mass media intervention). Included studies had to compare an eHealth intervention group to at least a control group or a non-eHealth intervention group.

Type of Studies

Articles of randomized controlled trials were eligible for inclusion. Pure qualitative assessments of the effectiveness of an intervention were not eligible.

Type of Outcomes

This review primarily focused on the following behavioral outcomes measured by either subjective or objective approaches:

(1) PA-related outcomes: energy expenditure, steps, time spent in moderate to vigorous PA (MVPA) and (2) healthy diet–related outcomes: energy intake, macronutrient composition (carbohydrate, protein, and lipids), core food group consumption (eg, all healthy components, vegetables, fruit, grain). The included articles should address both PA and healthy diet–related outcomes. In addition to the aforementioned primary outcomes, weight-related outcomes (eg, body weight, BMI, waist circumference, body fat, waist-to-hip ratio) were also considered as secondary outcomes.

Data Extraction

Data extraction was performed according to a study-created extraction tool; the main framework of the extraction tool was drawn from a previously published example [22]. Two authors conducted the initial data extraction, and a third reviewer was consulted if any discrepancies in data extraction were identified. The following information was extracted: (1) basic study characteristics including the first author, date of publication, and country of origin; (2) participant characteristics including sample size, age, gender ratio (female), disease type, and recruitment location; (3) intervention characteristics including delivery channel, intervention duration and intensity, underpinning theories, control group information, and detailed intervention content; (4) outcome measures including measurement of the outcomes and measuring points; and (5) main results including intervention completion ratio and converted effect size (standardized mean difference [SMD]).

Bias Assessments

Risk of bias was independently assessed by 2 authors according to the "Cochrane Collaboration's tool for assessing risk of bias" (selection, performance, detection, attrition, reporting, and other biases) [23]. Disagreement between the reviewers was resolved through mutual discussion until consensus was finally reached. In addition, publication bias was assessed using a nonparametric test based on the rank correlation between the estimated treatment effect and its variance [24].

Strategy for Data Synthesis and Meta-analysis

Results were pooled in the meta-analysis if the final values at postintervention were available. For the articles with continuous data, numbers of participants, mean scores, and SDs of the outcome variables were extracted to calculate the SMD: $(m_1-m_2)/\sqrt{[(s_1^2+ s_2^2) / 2]}$, where m is the mean and s is the SD.

For the articles with dichotomous data, the numbers of people in each category of both intervention and control groups were extracted to calculate the odds ratio (OR): N_a*N_d / N_b*N_c, where N_a is the number of adherents in the intervention group, N_b is the number of adherents in the control group, N_c is the number of nonadherents in the intervention group, and N_d is the number of nonadherents in the control group.

For the convenience of further calculation, the ORs were arithmetically converted to SMDs using a spreadsheet [25]. It should be noted that the effect size of some negative outcomes (eg, fat intake, BMI, unhealthy diet) were reverse coded into positive values.

For studies with multiple effect sizes in a particular scope of outcome (eg, PA-related outcomes, light PA, MVPA simultaneously), we used the weighted arithmetic averaging method to pool the effect sizes into a synthesized size [26]. Taking the study of Bantum et al [27] as an example, in this study there were 2 effect sizes regarding PA-related outcomes (SMD .16, light PA; SMD .38, MVPA). Since the sample sizes of these 2 effective sizes were equal (50% weighting each), the synthesized SMD for PA was (0.16 + 0.38)/2 = 0.27.

For the issues of missing data and statistics (eg, cases such as SDs and means not reported), we first resorted to statistical conversion (eg, convert 95% CI to SD). If it could not be statistically converted, we then contacted the authors directly for the datasets. If neither approach worked out, we excluded the unqualified studies listwise for the meta-analyses.

Finally, effect sizes were synthesized using a random effects model [28]. We adopted the random effects model because it allows inferences that generalize beyond the studies included in the specific meta-analysis [28]. A positive SMD reflects the between-group difference in favor of the eHealth MHBC intervention group over the control group (ie, increase in outcomes regarding PA, healthy diet, and healthy weight such as healthy BMI range). The pooled effect sizes are presented in forest plots that allow readers to see the information from the individual studies that went into the meta-analysis at a glance [29]. Heterogeneity was assessed using the recommended I^2 for Cochrane reviews [30]. Subgroup analyses were performed based on 5 binary variables: (1) gender (ratio of female participants ≥50% vs <50%), (2) age (participants' mean age ≥55 years vs <55 years), (3) intervention duration (≥24 weeks vs <24 weeks), (4) theory (whether the intervention was guided by theory), and (5) intervention channel (pure SMS or telephone vs internet- or web-based).

All data calculations (publication bias tests, effect size syntheses, heterogeneity tests, subgroup analyses) in this meta-analysis were conducted using SPSS 22.0 [31] with the syntaxes provided by Field and Gillett [28]. All the qualitative extractions (information extraction, risk of bias assessment) and figure productions were facilitated by Review Manager 5.3 [32].

Results

Study Characteristics

The initial systematic search yielded a total of 664 articles (see Figure 1). After duplicate deletion and screening of abstracts and full texts, 12 articles met the inclusion criteria, with 3 additional articles added by hand searches. In total, 15 studies met the inclusion criteria and did not meet the exclusion criteria targeting improvements in PA, healthy diet, and weight status among participants with NCDs. Due to the oversize issue, the main contents of each study are summarized in Multimedia Appendix 2. Of the 15 studies identified, 4 were conducted in the United States [26,33-35], 2 each in Canada [36,37] and India [38,39], and 1 each in China [18], Pakistan [40], Korea [41], the Netherlands [19], New Zealand [42], and the United Kingdom [43]. Notably, 1 study recruited its sample globally in Canada, the United Kingdom, and the United States [44].

Study participant numbers (intervention group and control group together) ranged from 59 [41] to 683 [37], with a mean sample size of 315 participants (SD 215 participants). The participants' average age ranged from 42.3 years [41] to 73.0 years [44]. For the gender ratio, 2 studies notably recruited predominantly or all (≥80%) male participants [38,42], while another 2 studies recruited female participants (>80%) [26,41]; 1 study did not report gender information [39]. Various NCD types were covered, among which, the top 3 addressed diseases were heart disease [18,34,37,42,43], cancer [19,26,41,44], and diabetes [33,38-40].

Various intervention channels and media were applied, including web sites or pages, telephone counselling, and SMS. Compared with the traditional SMS-only and telephone-only interventions, adding web-based materials in interventions were more prevalent (8/15, 53%). Notably, many studies adopted a mixed-channel intervention, such as combining a web-based intervention with SMS reminders and offline peer support or group meetings [26,33] or combining web-based intervention material with SMS or telephone reminders [18,19,41]. The intervention durations ranged from 6 weeks [26] to 1 year [39,43]. Of the included studies, 9 designed interventions of no shorter than 24 weeks, whereas 6 studies designed interventions shorter than 24 weeks. The majority (9/15, 60%) of the included interventions were designed based on a particular theory, and among the interventions with a theoretical backdrop, the Transtheoretical Model (6/15, 40%) and Social Cognitive Theory (4/15, 27%) were the top 2 frequently supporting theories.

All 15 studies evaluated both PA and healthy diet–related outcomes. However, 2 studies [34,35] did not provide SDs of both the PA and healthy diet postintervention outcomes. We failed to obtain these missing results from the corresponding authors via email, thus making these studies ineligible for the meta-analysis. In addition, 4 of the 15 studies assessed weight-related outcomes [33,38-40]. The intervention completion rate (mean numbers of participants who finished the entire intervention divided by the number of participants at baseline) ranged from 70.9% [39] to 96.7% [19].

Bias and Heterogeneity Assessments

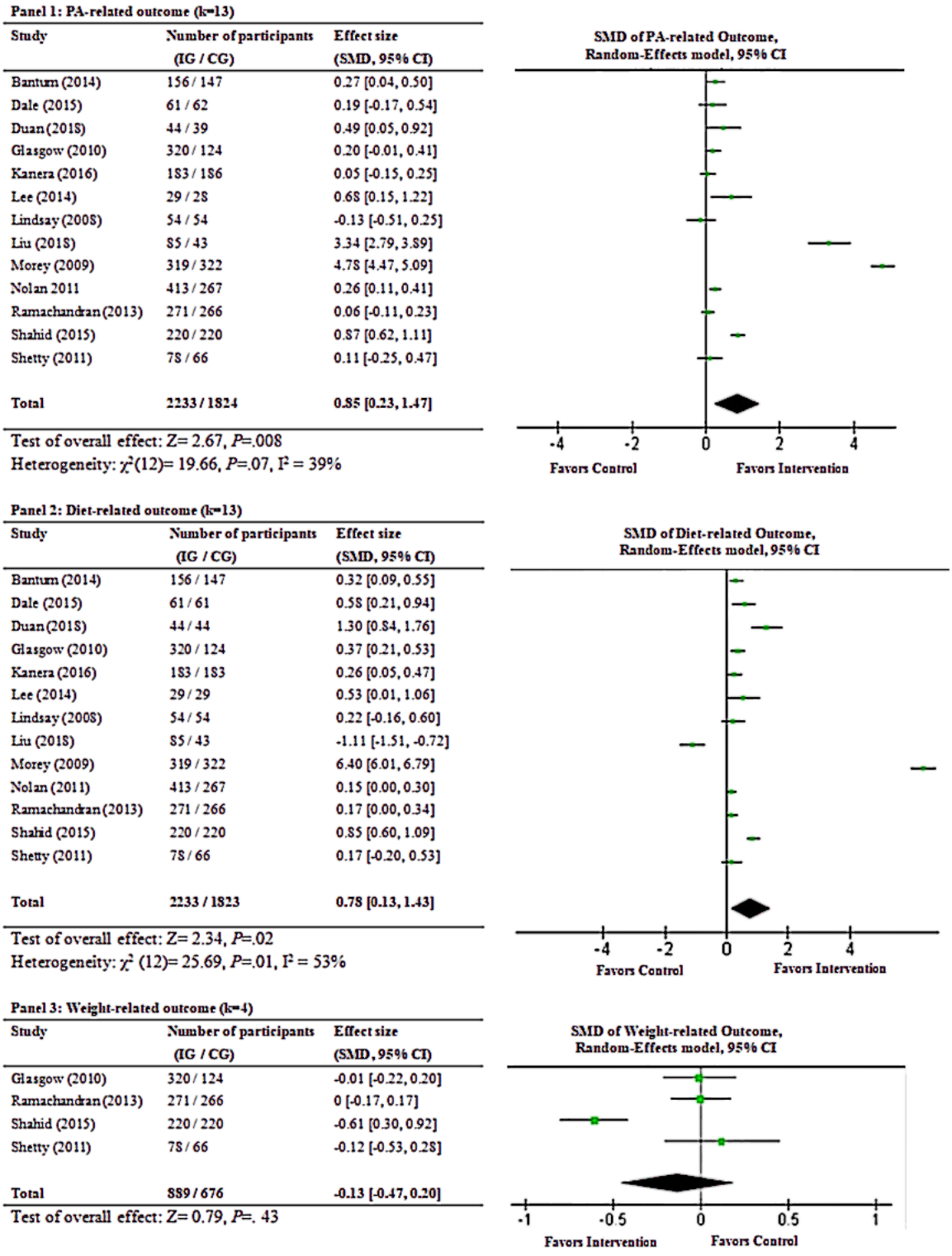
As can be seen in Figure 2, the risk of bias assessment indicated that the included articles were quite high in quality. For the publication bias examination, the statistical results showed no potential publication bias regarding the 3 intervention outcomes of PA (Kendall tau=.21, *P*=.33), healthy diet (Kendall tau=.26, *P*=.22), and weight (Kendall tau=.33, *P*=.50).

Figure 2. Risk of bias in individual studies (k = 15).

Intervention Effectiveness

In the following sections, the synthesized results regarding PA-related, diet-related, and weight-related outcomes are introduced individually. The main results are visualized in the forest plot in Figure 3.

Figure 3. Meta-analysis of outcomes related to physical activity (PA; panel 1), diet (panel 2), and weight (panel 3). CG: control group; IG: intervention group.



PA-Related Outcomes

The synthesized effect size from the 13 studies demonstrated significant differences in total PA between intervention and control groups at postintervention (SMD 0.85, Z= 2.67, 95% CI 0.23 to 1.47, *P*=.008; see panel 1 in Figure 3). According to the rule of thumb [45], an SMD of 0.85 should be considered a large effect size. The heterogeneity test did not show significance among PA-related outcomes (χ^2_{12}=19.66, *P*=.07, I^2=39%). For the subgroup analyses, gender (χ^2_1=0.44, *P*=.51), age (χ^2_1=1.78, *P*=.18), intervention duration (χ^2_1=0.98, *P*=.32), and intervention channel (χ^2_1=0.75, *P*=.39) were all not significantly related to the intervention effectiveness. Noticeably, whether the intervention was based on theory (χ^2_1=2.41, *P*=.12) marginally approached the subgroup

significant difference. Specifically, interventions based on theories (SMD 1.22, Z=2.13, 95% CI 0.10 to 2.34, *P*=.03) achieved better effectiveness than nontheory-based interventions (SMD 0.30, Z=1.83, 95% CI –0.02 to 0.61, *P*=.07).

Healthy Diet–Related Outcomes

In terms of the remaining 13 studies, significant differences in healthy diet behaviors were demonstrated between intervention and control groups at postintervention (SMD 0.78, Z=2.34, 95% CI 0.13 to 1.43, *P*=.02; see panel 2 in Figure 3). According to the rule of thumb [45], an SMD of 0.78 should be considered a medium effect size. Concerning the heterogeneity tests, significant heterogeneity was revealed among healthy diet–related outcomes (χ^2_{12}=25.69, *P*=.01, I^2=53%). For the subgroup analyses, gender χ^2_1=2.19, *P*=.14), age (χ^2_1=.46, *P*=.50), intervention duration (χ^2_1=.21, *P*=.65), and whether the intervention was based on theory (χ^2_1=1.28, *P*=.26) were not significantly related to the intervention effectiveness. However, the intervention channel was marginally related to the intervention effectiveness (χ^2_1=2.59, *P*=.10). Specifically, interventions with traditional SMS or telephone counselling (SMD 1.54, Z=2.05, 95% CI 0.07 to 3.01, *P*=.04) achieved better effectiveness than web-based interventions (SMD 0.30, Z=1.70, 95% CI –0.05 to 0.64, *P*=.09).

Weight-Related Outcomes

There were only 4 studies examining weight-related outcomes. Results did not show a significant difference in weight status between intervention and control groups at postintervention (SMD –0.13, Z=0.79, 95% CI –0.47 to 0.20, *P*=.43; see panel 3 in Figure 3). Insignificant heterogeneity was shown among weight-related outcomes (χ^2_3=2.82, *P*=.42). I^2 could not be calculated due to the small number of studies. Since subgroup analyses is not recommended when any subgroup has <4 studies [46], we did not conduct the subgroup analysis for the weight-related outcomes.

Discussion

Principal Findings

This review systematically identified 15 studies that investigated the effectiveness of MHBC eHealth interventions aimed at improving PA-, healthy diet–, and weight-related outcomes among people with NCDs. The results showed that the MHBC eHealth interventions significantly promoted both PA and healthy diet outcomes among people with NCDs. However, the results did not show significant intervention effectiveness regarding weight changes. These results are in partial agreement with the review findings of Amireault et al [47], which showed MHBC interventions regarding PA- and healthy diet–related behaviors could significantly improve PA behaviors among cancer survivors. Yet, the findings from our review differ from the findings of the review by Alageel et al [48], which demonstrated the majority of MHBC interventions for patients with cardiovascular diseases could not achieve significant changes in either PA behavior or fruit and vegetable consumption. Regarding effect sizes, MHBC eHealth interventions achieved a large effect size in terms of PA (SMD 0.85) and healthy diet (SMD 0.78) behaviors [45]. The effect sizes from our meta-analysis are larger than the effect sizes of 2 meta-analyses synthesizing eHealth interventions on PA behavior in older people (SMD 0.79) [49] and fruit and vegetable intake in a healthy population (SMD 0.26) [50]. This might be due to the fact that we targeted people with NCDs who are already impacted by diseases. Thus, the populations in our meta-analyses might be more motivated and willing than other healthy populations to change their health behaviors, thus reaping more positive intervention effects [51].

Though all the included intervention studies adopted an MHBC eHealth approach, there was still a high variability in participants (eg, cultural backgrounds, age), intervention characteristics (eg, intervention channel, content, duration), and outcome measurements. This indicates that, as an emerging intervention paradigm, MHBC eHealth interventions are in the exploratory phase and do not have relatively well-acknowledged intervention guidelines or standards such as CONSORT [52].

Despite this high variability, we still found notable trends. First, the eHealth intervention channel has clearly changed from traditional SMS-based or telephone counselling to modern, multimedia, web-based or smartphone-based interventions. Since 2015, except for 1 article using a traditional telephone-based intervention [34], all interventions were web-based. With the convenience of an internet connection and the popularity of smartphone usage in daily life, it has been predicted that future interventions via smartphone apps are promising for better promotion of a healthy lifestyle among people with NCDs [13].

Second, the intervention contents of most included studies mainly focused on health behavior education and counselling without substantial behavioral tutorials; this might be due to the present limitations of the eHealth intervention channel. The commonly used intervention paradigm was to select a particular health behavior change theory as the framework and further promote the effective elements (e.g., motivation, planning, and self-regulation) of the chosen theory. Such commonly used approaches did achieve significant medium-sized effectiveness of the intervention (SMD around 0.8) in changing the weight-related health behaviors of PA and healthy diet. To further increase the intervention effectiveness, a dual-process approach (ie, focusing on both conscious and nonconscious processes of behavior change) [53,54] and a social-ecological approach (ie, involving policy-level, environmental, and personal factors) may be prudent [55,56].

Third, regarding the outcome measurements, the majority of studies used self-report measures. Self-report is a feasible, economic, and time-saving approach for data collection [57], but it also has limitations of high subjectivity and low accuracy caused by social desirability and reporting bias. With the advancement of technology, more objective approaches of data collection regarding PA (eg, geo-information, system-based recording; wearable device–based recording), healthy diet (eg, food photography, computer-assisted recall), and weight (eg, electronic scale data collection) are recommended to improve the accuracy during data collection [58].

Fourth, follow-up analyses were generally lacking in the intervention designs. Most of the included studies measured outcomes twice (preintervention baseline and postintervention). Long-term and maintenance effects of MHBC eHealth interventions on both PA and healthy diet behaviors among people with NCDs were not therefore validated. Given that the MHBC eHealth intervention is in its infancy, these first studies were mainly to examine whether eHealth interventions with PA and healthy diet could be effective. As advocated [6,13], the next stage in this research is to explore how and under what conditions these initial changes can be maintained by adding longer follow-up designs.

Regarding the intervention effectiveness, we adopted analytical methods proposed in previous relevant systematic review articles to pool the effect sizes by different outcomes [47,59]. This approach can effectively reduce the difficulty of data processing and improve the clarity of results presentation. On the other hand, this approach somewhat ignored the overall effect of a particular intervention program. Our results only indicated that MHBC eHealth interventions targeting people with NCDs, on average, can significantly promote PA and healthy diet; it has yet to be confirmed that MHBC eHealth interventions can successfully promote an entire pattern of weight-related health behaviors among people with NCDs. In the future, as the volume of MHBC studies increases, a well-acknowledged index to indicate the overall effect size of MHBC interventions is expected to be developed for better evaluation of intervention effects.

Interestingly, the post hoc subgroup analyses did not find any significant moderators influencing intervention effectiveness. However, the results did reveal some potential moderators (ie, whether the intervention was based on theory, intervention channel) reaching marginal significance, which is noteworthy given the small sample of studies. Interventions based on theories achieved significant effectiveness for promoting PA behaviors, while those without any theoretical basis did not. The results are consistent with previous meta-analyses regarding using theory in health behavior promotion [60-63]. And interestingly, interventions with traditional intervention media achieved better effectiveness than web-based interventions for promoting healthy diet behaviors. The reason might be that some of the included web-based interventions were lacking direct virtual communication or periodic reminders, which might potentially hinder the intervention effectiveness [36,43]. This inference could be supported by previous research that web-based interventions can achieve a desirable effect only when adding additional methods of communication with participants, especially the use of SMS or text messages [61].

Also interestingly, intervention duration did not have a significant impact on intervention effectiveness for either PA or healthy diet behaviors; a possible explanation might be that the intervention frequency and intervention sequence were not considered [64]. In summary, we recommend future MHBC eHealth interventions combine web-based intervention material with traditional periodic calling or reminders through virtual contact and pay more attention to intervention quality and sequence arrangement among multiple health behaviors for achieving better intervention effectiveness.

Limitations

Three limitations of this review should be acknowledged. First, despite our best efforts to conduct a thorough literature search in the limited databases, it may still have resulted in the omission of suitable topics or related studies due to not including key terms or research outside the time span that was searched. Second, bias might be present in some included studies because they lacked a registered protocol, used inappropriate statistical methods, and had missing information [39,40,43], which therefore suggests additional caution should be taken in interpreting the findings from these trials and the pooled effect sizes. Third, there was a high degree of heterogeneity (ie, participant characteristics, intervention types and lengths, and outcome measurements), and the small number of studies could lead to cautious interpretation of the synthesized results.

Conclusions

To the best of our knowledge, this review is the only study that has attempted to synthesize the literature regarding the effectiveness of MHBC eHealth interventions on PA, healthy diet, and weight for people with NCDs. Such MHBC eHealth studies have emerged in recent years as a new trend in aftercare rehabilitation settings. The current review significantly contributes to the eHealth- and NCD-related literature by identifying research priorities and providing preliminary evidence for clinical decision making. This review indicates that MHBC eHealth interventions have obtained preliminary success in promoting PA and healthy diet behaviors among people with NCDs. The identification of critical intervention characteristics such as being theory-based and adding communication elements to web-based intervention material are essential for maximizing the effects of MHBC eHealth interventions in promoting weight-related behaviors among people with NCDs. Based on this review study, it is expected that further investigations will make recommended improvements on the intervention design in order to ultimately enhance the well-being of people with NCDs.

Acknowledgments

We would like to thank Professor Martin Hagger for his helpful instruction and advice on the statistics for the meta-analysis. This research was not supported by a specific grant from any funding agency.

Authors' Contributions

YD, BS, and RR designed the study and revised the manuscript. BS and YD drafted the manuscript. BS, WL, GD, MY, and YD screened and selected the studies. BS, GD, MY, and YD extracted the data. BS, GD, and YD analyzed the data. All the authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The search strategies of the systematic review.
[DOCX File , 16 KB - jmir_v23i2e23786_app1.docx]

Multimedia Appendix 2
Detailed records of eligible studies (k= 15).
[DOC File , 78 KB - jmir_v23i2e23786_app2.doc]

References

1. Noncommunicable diseases. World Health Organization. 2018 Jun. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2021-02-06]
2. Bloom D, Chen S, Kuhn M, McGovern M, Oxley L, Prettner K. The economic burden of chronic diseases: Estimates and projections for China, Japan, and South Korea. The Journal of the Economics of Ageing 2020 Oct;17:100163 [FREE Full text] [doi: 10.1016/j.jeoa.2018.09.002]
3. Jakovljevic MB, Milovanovic O. Growing Burden of Non-Communicable Diseases in the Emerging Health Markets: The Case of BRICS. Front Public Health 2015;3:65 [FREE Full text] [doi: 10.3389/fpubh.2015.00065] [Medline: 25954740]
4. Broers ER, Kop WJ, Denollet J, Widdershoven J, Wetzels M, Ayoola I, et al. A Personalized eHealth Intervention for Lifestyle Changes in Patients With Cardiovascular Disease: Randomized Controlled Trial. J Med Internet Res 2020 May 22;22(5):e14570 [FREE Full text] [doi: 10.2196/14570] [Medline: 32441658]
5. Prochaska JO. Multiple Health Behavior Research represents the future of preventive medicine. Prev Med 2008 Mar;46(3):281-285. [doi: 10.1016/j.ypmed.2008.01.015] [Medline: 18319100]
6. Geller K, Lippke S, Nigg CR. Future directions of multiple behavior change research. J Behav Med 2017 Feb;40(1):194-202. [doi: 10.1007/s10865-016-9809-8] [Medline: 27785652]
7. Nigg CR, Long CR. A systematic review of single health behavior change interventions vs. multiple health behavior change interventions among older adults. Transl Behav Med 2012 Jun;2(2):163-179 [FREE Full text] [doi: 10.1007/s13142-012-0130-y] [Medline: 24073109]
8. Beaglehole R, Bonita R, Horton R, Adams C, Alleyne G, Asaria P, et al. Priority actions for the non-communicable disease crisis. The Lancet 2011 Apr 23;377(9775):1438-1447. [doi: 10.1016/S0140-6736(11)60393-0] [Medline: 21474174]
9. Prochaska JJ, Prochaska JO. A Review of Multiple Health Behavior Change Interventions for Primary Prevention. Am J Lifestyle Med 2011 May;5(3):208-221 [FREE Full text] [doi: 10.1177/1559827610391883] [Medline: 24358034]
10. Rhodes RE, Quinlan A, Mistry CD. Do other goals influence physical activity? A systematic review examining the relationship between other goals and physical activity behavior. Prev Med 2016 Oct;91:306-317. [doi: 10.1016/j.ypmed.2016.08.033] [Medline: 27568235]
11. van Dam RM, Li T, Spiegelman D, Franco OH, Hu FB. Combined impact of lifestyle factors on mortality: prospective cohort study in US women. BMJ 2008 Sep 16;337:a1440 [FREE Full text] [doi: 10.1136/bmj.a1440] [Medline: 18796495]
12. Buis LR, Dawood K, Kadri R, Dawood R, Richardson CR, Djuric Z, et al. Improving Blood Pressure Among African Americans With Hypertension Using a Mobile Health Approach (the MI-BP App): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Jan 25;8(1):e12601 [FREE Full text] [doi: 10.2196/12601] [Medline: 30681965]
13. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, Present, and Future of eHealth and mHealth Research to Improve Physical Activity and Dietary Behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
14. Boogerd EA, Arts T, Engelen LJ, van de Belt TH. "What Is eHealth": Time for An Update? JMIR Res Protoc 2015 Mar 12;4(1):e29 [FREE Full text] [doi: 10.2196/resprot.4065] [Medline: 25768939]
15. Murray E, Khadjesari Z, White IR, Kalaitzaki E, Godfrey C, McCambridge J, et al. Methodological challenges in online trials. J Med Internet Res 2009 Apr 03;11(2):e9 [FREE Full text] [doi: 10.2196/jmir.1052] [Medline: 19403465]
16. Haberlin C, O'Dwyer T, Mockler D, Moran J, O'Donnell DM, Broderick J. The use of eHealth to promote physical activity in cancer survivors: a systematic review. Support Care Cancer 2018 Oct 16;26(10):3323-3336. [doi: 10.1007/s00520-018-4305-z] [Medline: 29909476]

17. Rollo ME, Aguiar EJ, Williams RL, Wynne K, Kriss M, Callister R, et al. eHealth technologies to support nutrition and physical activity behaviors in diabetes self-management. Diabetes Metab Syndr Obes 2016;9:381-390 [FREE Full text] [doi: 10.2147/DMSO.S95247] [Medline: 27853384]
18. Duan YP, Liang W, Guo L, Wienert J, Si GY, Lippke S. Evaluation of a Web-Based Intervention for Multiple Health Behavior Changes in Patients With Coronary Heart Disease in Home-Based Rehabilitation: Pilot Randomized Controlled Trial. J Med Internet Res 2018 Nov 19;20(11):e12052 [FREE Full text] [doi: 10.2196/12052] [Medline: 30455167]
19. Kanera IM, Bolman CAW, Willems RA, Mesters I, Lechner L. Lifestyle-related effects of the web-based Kanker Nazorg Wijzer (Cancer Aftercare Guide) intervention for cancer survivors: a randomized controlled trial. J Cancer Surviv 2016 Oct 17;10(5):883-897 [FREE Full text] [doi: 10.1007/s11764-016-0535-6] [Medline: 26984534]
20. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
21. Su X, Dai H, Zhang F, Chen K, Li H, Wu X, et al. Research on the Application of "Internet + medical treatment" in Diabetes Hospital - community - family Management Model. Hospital Management Forum 2018;35(4):74-77.
22. Müller AM, Alley S, Schoeppe S, Vandelanotte C. Int J Behav Nutr Phys Act 2016 Oct 10;13(1):109 [FREE Full text] [doi: 10.1186/s12966-016-0434-2] [Medline: 27724911]
23. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
24. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994 Dec;50(4):1088-1101. [Medline: 7786990]
25. DeCoster J. Converting effect sizes Microsoft Excel Spreadsheets. Stat-Help. 2012. URL: http://www.stat-help.com/spreadsheets.html [accessed 2021-02-06]
26. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. Hoboken, NJ: John Wiley & Sons, Ltd; Mar 11, 2009.
27. Bantum EO, Albright CL, White KK, Berenberg JL, Layi G, Ritter PL, et al. Surviving and thriving with cancer using a Web-based health behavior change intervention: randomized controlled trial. J Med Internet Res 2014 Feb 24;16(2):e54 [FREE Full text] [doi: 10.2196/jmir.3020] [Medline: 24566820]
28. Field AP, Gillett R. How to do a meta-analysis. Br J Math Stat Psychol 2010 Nov;63(Pt 3):665-694. [doi: 10.1348/000711010X502733] [Medline: 20497626]
29. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. BMJ 2001 Jun 16;322(7300):1479-1480 [FREE Full text] [doi: 10.1136/bmj.322.7300.1479] [Medline: 11408310]
30. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions, Version 5.0.0. 2008. URL: https://training.cochrane.org/handbook/archive/v5.0.0/ [accessed 2021-02-06]
31. SPSS Statistics for Windows, Version 22. IBM. 2013. URL: https://www.ibm.com/support/pages/spss-statistics-220-available-download [accessed 2021-02-06]
32. Cochrane Review Manager (RevMan). 2014. URL: revman.cochrane.org [accessed 2021-02-06]
33. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Outcomes of minimal and moderate support versions of an internet-based diabetes self-management support program. J Gen Intern Med 2010 Dec;25(12):1315-1322 [FREE Full text] [doi: 10.1007/s11606-010-1480-0] [Medline: 20714820]
34. Swoboda CM, Miller CK, Wills CE. Setting Single or Multiple Goals for Diet and Physical Activity Behaviors Improves Cardiovascular Disease Risk Factors in Adults With Type 2 Diabetes: A Pragmatic Pilot Randomized Trial. Diabetes Educ 2016 Aug;42(4):429-443. [doi: 10.1177/0145721716650043] [Medline: 27194048]
35. Migneault JP, Dedier JJ, Wright JA, Heeren T, Campbell MK, Morisky DE, et al. A culturally adapted telecommunication system to improve physical activity, diet quality, and medication adherence among hypertensive African-Americans: a randomized controlled trial. Ann Behav Med 2012 Feb;43(1):62-73. [doi: 10.1007/s12160-011-9319-4] [Medline: 22246660]
36. Liu S, Brooks D, Thomas SG, Eysenbach G, Nolan RP. Effectiveness of User- and Expert-Driven Web-based Hypertension Programs: an RCT. Am J Prev Med 2018 Apr;54(4):576-583. [doi: 10.1016/j.amepre.2018.01.009] [Medline: 29456025]
37. Nolan RP, Upshur REG, Lynn H, Crichton T, Rukholm E, Stewart DE, et al. Therapeutic benefit of preventive telehealth counseling in the Community Outreach Heart Health and Risk Reduction Trial. Am J Cardiol 2011 Mar 1;107(5):690-696. [doi: 10.1016/j.amjcard.2010.10.050] [Medline: 21215382]
38. Ramachandran A, Snehalatha C, Ram J, Selvam S, Simon M, Nanditha A, et al. Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: a prospective, parallel-group, randomised controlled trial. The Lancet Diabetes & Endocrinology 2013 Nov;1(3):191-198. [doi: 10.1016/S2213-8587(13)70067-6] [Medline: 24622367]
39. Shetty AS, Chamukuttan S, Nanditha A, Raj RKC, Ramachandran A. Reinforcement of adherence to prescription recommendations in Asian Indian diabetes patients using short message service (SMS)--a pilot study. J Assoc Physicians India 2011 Nov;59:711-714. [Medline: 22616337]

40. Shahid M, Mahar SA, Shaikh S, Shaikh Z. Mobile phone intervention to improve diabetes care in rural areas of Pakistan: a randomized controlled trial. J Coll Physicians Surg Pak 2015 Mar;25(3):166-171. [doi: 10.1016/S1530-891X(20)42530-3] [Medline: 25772954]
41. Lee MK, Yun YH, Park H, Lee ES, Jung KH, Noh D. A Web-based self-management exercise and diet intervention for breast cancer survivors: pilot randomized controlled trial. Int J Nurs Stud 2014 Dec;51(12):1557-1567. [doi: 10.1016/j.ijnurstu.2014.04.012] [Medline: 24856854]
42. Pfaeffli Dale L, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text Message and Internet Support for Coronary Heart Disease Self-Management: Results From the Text4Heart Randomized Controlled Trial. J Med Internet Res 2015 Oct 21;17(10):e237 [FREE Full text] [doi: 10.2196/jmir.4944] [Medline: 26490012]
43. Lindsay S, Bellaby P, Smith S, Baker R. Enabling healthy choices: is ICT the highway to health improvement? Health (London) 2008 Jul;12(3):313-331. [doi: 10.1177/1363459308090051] [Medline: 18579630]
44. Morey MC, Snyder DC, Sloane R, Cohen HJ, Peterson B, Hartman TJ, et al. Effects of home-based diet and exercise on functional outcomes among older, overweight long-term cancer survivors: RENEW: a randomized controlled trial. JAMA 2009 May 13;301(18):1883-1891 [FREE Full text] [doi: 10.1001/jama.2009.643] [Medline: 19436015]
45. Sawilowsky S. New Effect Size Rules of Thumb. J. Mod. App. Stat. Meth 2009 Nov 01;8(2):597-599 [FREE Full text] [doi: 10.22237/jmasm/1257035100]
46. Fu R, Gartlehner G, Grant M, Shamliyan T, Sedrakyan A, Wilt TJ, et al. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol 2011 Nov;64(11):1187-1197. [doi: 10.1016/j.jclinepi.2010.08.010] [Medline: 21477993]
47. Amireault S, Fong AJ, Sabiston CM. Promoting Healthy Eating and Physical Activity Behaviors: A Systematic Review of Multiple Health Behavior Change Interventions Among Cancer Survivors. Am J Lifestyle Med 2018;12(3):184-199 [FREE Full text] [doi: 10.1177/1559827616661490] [Medline: 30202391]
48. Alageel S, Gulliford MC, McDermott L, Wright AJ. Multiple health behaviour change interventions for primary prevention of cardiovascular disease in primary care: systematic review and meta-analysis. BMJ Open 2017 Jun 15;7(6):e015375 [FREE Full text] [doi: 10.1136/bmjopen-2016-015375] [Medline: 28619779]
49. Kwan RYC, Salihu D, Lee PH, Tse M, Cheung DSK, Roopsawang I, et al. The effect of e-health interventions promoting physical activity in older people: a systematic review and meta-analysis. Eur Rev Aging Phys Act 2020;17:7 [FREE Full text] [doi: 10.1186/s11556-020-00239-5] [Medline: 32336996]
50. Rodriguez Rocha NP, Kim H. eHealth Interventions for Fruit and Vegetable Intake: A Meta-Analysis of Effectiveness. Health Educ Behav 2019 Dec;46(6):947-959. [doi: 10.1177/1090198119859396] [Medline: 31347403]
51. Sin DYE, Guo X, Yong DWW, Qiu TY, Moey PKS, Falk M, et al. Assessment of willingness to Tele-monitoring interventions in patients with type 2 diabetes and/or hypertension in the public primary healthcare setting. BMC Med Inform Decis Mak 2020 Jan 28;20(1):11 [FREE Full text] [doi: 10.1186/s12911-020-1024-4] [Medline: 31992288]
52. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c332 [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]
53. Sheeran P, Gollwitzer PM, Bargh JA. Nonconscious processes and health. Health Psychol 2013 May;32(5):460-473. [doi: 10.1037/a0029203] [Medline: 22888816]
54. Hagger MS. Non-conscious processes and dual-process theories in health psychology. Health Psychol Rev 2016 Dec;10(4):375-380. [doi: 10.1080/17437199.2016.1244647] [Medline: 27718880]
55. Rhodes RE, McEwan D, Rebar AL. Theories of physical activity behaviour change: A history and synthesis of approaches. Psychology of Sport and Exercise 2019 May;42:100-109. [doi: 10.1016/j.psychsport.2018.11.010]
56. Kang H, Han J, Kwon GH. An Ecological Approach to Smart Homes for Health Care Services: Conceptual Framework of a Smart Servicescape Wheel. JMIR Mhealth Uhealth 2019 Feb 08;7(2):e12425 [FREE Full text] [doi: 10.2196/12425] [Medline: 30735145]
57. Annett J. Subjective rating scales: science or art? Ergonomics 2002 Nov 15;45(14):966-987. [doi: 10.1080/00140130210166951] [Medline: 12569049]
58. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices. Med Sci Sports Exerc 2016 Mar 17;48(3):457-465 [FREE Full text] [doi: 10.1249/MSS.0000000000000778] [Medline: 26484953]
59. Champion KE, Newton NC, Spring B, Wafford QE, Parmenter BJ, Teesson M. A systematic review of school-based eHealth interventions targeting alcohol use, smoking, physical inactivity, diet, sedentary behaviour and sleep among adolescents: a review protocol. Syst Rev 2017 Dec 06;6(1):246 [FREE Full text] [doi: 10.1186/s13643-017-0645-x] [Medline: 29208040]
60. Prestwich A, Sniehotta FF, Whittington C, Dombrowski SU, Rogers L, Michie S. Does theory influence the effectiveness of health behavior interventions? Meta-analysis. Health Psychol 2014 May;33(5):465-474. [doi: 10.1037/a0032853] [Medline: 23730717]
61. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

62. McEwan D, Beauchamp MR, Kouvousis C, Ray CM, Wyrough A, Rhodes RE. Examining the active ingredients of physical activity interventions underpinned by theory versus no stated theory: a meta-analysis. Health Psychol Rev 2019 Mar;13(1):1-17. [doi: 10.1080/17437199.2018.1547120] [Medline: 30412685]
63. Hagger MS, Weed M. DEBATE: Do interventions based on behavioral theory work in the real world? Int J Behav Nutr Phys Act 2019 Apr 25;16(1):36 [FREE Full text] [doi: 10.1186/s12966-019-0795-4] [Medline: 31023328]
64. James E, Freund M, Booth A, Duncan MJ, Johnson N, Short CE, et al. Comparative efficacy of simultaneous versus sequential multiple health behavior change interventions among adults: A systematic review of randomised trials. Prev Med 2016 Aug;89:211-223. [doi: 10.1016/j.ypmed.2016.06.012] [Medline: 27311332]

Abbreviations

MHBC: multiple health behavior change
MVPA: moderate to vigorous physical activity
NCD: noncommunicable disease
OR: odds ratio
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SMD: standardized mean difference

Edited by G Eysenbach, R Kukafka; submitted 23.08.20; peer-reviewed by J Alvarez Pitti, AS Aslam, B Lai; comments to author 01.10.20; revised version received 09.10.20; accepted 24.01.21; published 22.02.21.

Please cite as:
Duan Y, Shang B, Liang W, Du G, Yang M, Rhodes RE
Effects of eHealth-Based Multiple Health Behavior Change Interventions on Physical Activity, Healthy Diet, and Weight in People With Noncommunicable Diseases: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(2):e23786
URL: https://www.jmir.org/2021/2/e23786
doi:10.2196/23786
PMID:33616534

Review

An Education Framework for Effective Implementation of a Health Information System: Scoping Review

Tharshini Jeyakumar[1,2], BSc, MHI; Sarah McClure[1], CTDP; Mandy Lowe[1,3], MSc, OT Reg (Ont); Brian Hodges[1,4,5], PhD, MD, FRCPC; Katharine Fur[1], CTDP; Mariquita Javier-Brozo[1], BASc, RD; Maria Tassone[1,6], BScPT, MSc; Melanie Anderson[1], BAH, MLIS; Tim Tripp[1], BSc, MLIS; David Wiljer[1,2,4,7], PhD

[1]University Health Network, Toronto, ON, Canada

[2]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

[3]Department of Occupational Science and Occupational Therapy, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[4]Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[5]Wilson Centre, Toronto, ON, Canada

[6]Department of Physical Therapy, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[7]Office of Education, Centre for Addiction and Mental Health, Toronto, ON, Canada

Corresponding Author:
David Wiljer, PhD
University Health Network
190 Elizabeth Street, R. Fraser Elliot Building RFE 3S-441
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 340 4800 ext 6322
Email: David.wiljer@uhn.ca

Abstract

Background: To optimize their use of a new Health Information System (HIS), supporting health care providers require effective HIS education. Failure to provide this education can significantly hinder an organization's HIS implementation and sustainability efforts.

Objective: The aim of this review is to understand the most effective educational strategies and approaches to enable health care providers to optimally use an HIS.

Methods: Ovid MEDLINE, Ovid Embase, EBSCO Cumulative Index to Nursing and Allied Health Literature, and EBSCO Education Resources Information Center were searched to identify relevant papers. Relevant studies were systematically reviewed and analyzed using a qualitative thematic analysis approach.

Results: Of the 3539 studies screened, 17 were included for data extraction. The literature on the most effective approaches to enable health care providers to optimally use an HIS emphasized the importance of investing in engaging and understanding learners in the clinical context, maximizing the transfer of learning to care, and designing continuous and agile evaluation to meet the emerging demands of the clinical environment.

Conclusions: This review supports the advancement of a new HIS learning framework that organizational leaders and educators can use to guide HIS education design and development. Future research should examine how this framework can be translated into practice.

(J Med Internet Res 2021;23(2):e24691) doi:10.2196/24691

KEYWORDS

health information system; health care providers; education; learning; patient care

Introduction

Background

Health Information Systems (HISs) have been proposed as one solution in a multipronged organizational approach to transforming the quality of care delivered, increasing patient safety, and reducing health care costs [1]. An HIS is defined as a system designed to integrate data collection, processing, and reporting and the use of health information to influence policy making and improve health service effectiveness and efficiency [2]. This enables the facilitation of health information sharing among multiple authorized custodians across the health care continuum in support of clinical efficiency and optimal quality care [3]. Challenges to HIS implementation can result from staff and care providers being unfamiliar with the system features and functionalities, thus encumbering the ability of staff and providers to use HIS in their work environment most effectively. Given the rapid pace of adoption of digital HISs globally and the skills needed to effectively use HISs in practice, it is crucial to support and educate health care providers at all levels, across all areas of the health system, and as close to implementation as possible, on how the HIS system can be leveraged to improve clinical practice. Simply teaching the system is not sufficient to successfully enable learners (entry-to-practice and practicing health care staff) to use new technology.

As health care organizations respond to budgetary, regulatory, and societal pressures to implement an HIS, care providers are confronted with an ever-increasing technology-enabled care environment [4]. Paradoxically, although the implementation of HISs is meant to create efficiencies, their widespread diffusion and accompanying complexity have been associated with a growing recognition of clinician dissatisfaction and burnout [5]. However, health care providers see education as integral to use technology successfully [4]. On the basis of the model of skill acquisition, Bredfeldt et al [1] determined that educating novices entails more than just imparting knowledge. Further, it is essential to provide education that is pertinent to address learning needs and provide an opportunity to learn the nature of the real setting and its associated variability [1]. It is also imperative to acknowledge learners' expertise and engage them throughout all phases in the development of the educational program [6]. This may be especially true for health sciences students and trainees who may have deeper knowledge of using various technologies, thereby having the potential to play a role of educator or facilitator with supervisors in practice. Furthermore, education may enable care providers and staff to understand the concepts underlying the different HIS tasks as the emerging needs of their clinical practice continuously evolve rather than simply learning the features of an HIS. HIS education can also influence user adoption and the ability of health care providers and staff to effectively use the technology [4]. On the basis of learners' feedback in the study, McAlearney et al [7] noted that staff who received excellent education and hands-on experience with an information technology system adopted realistic expectations and achieved a sense of control within the HIS environment. Adequate education and support, technology literacy, and overall competencies of health providers were identified as critical factors in HIS implementation.

Many information technology users encounter a steep learning curve at the initial stages of implementation and can take several years to become an expert in the features and functions of a system [8]. McLean et al [8] suggested that users' attitudes toward system use in the initial stages could be leveraged to gain valuable insights into how the system will be used in the later stages. Thus, education remains a critical component in understanding the benefits of HISs and attaining value-added use, particularly during the adoption or early stages of implementation [8,9]. Furthermore, to enable ongoing learning in using an HIS adeptly, education strategies will need to be evaluated and refined over time, as education will need to continuously evolve to meet users' needs and comfort level with the HIS [8,9].

In several HIS implementation projects, inadequate HIS education has led to challenges in the adoption and suboptimal use of the system. An example is an electronic health record (EHR) implementation project at Cedars Sinai Hospital, California, where a dearth of staff education contributed to poor adjustment to the new system and created a sense of fear and apprehension [10,11]. Consequently, this lack of education threatened staff autonomy and eventually contributed to the rejection of the HIS and project failure [10,11].

Objectives

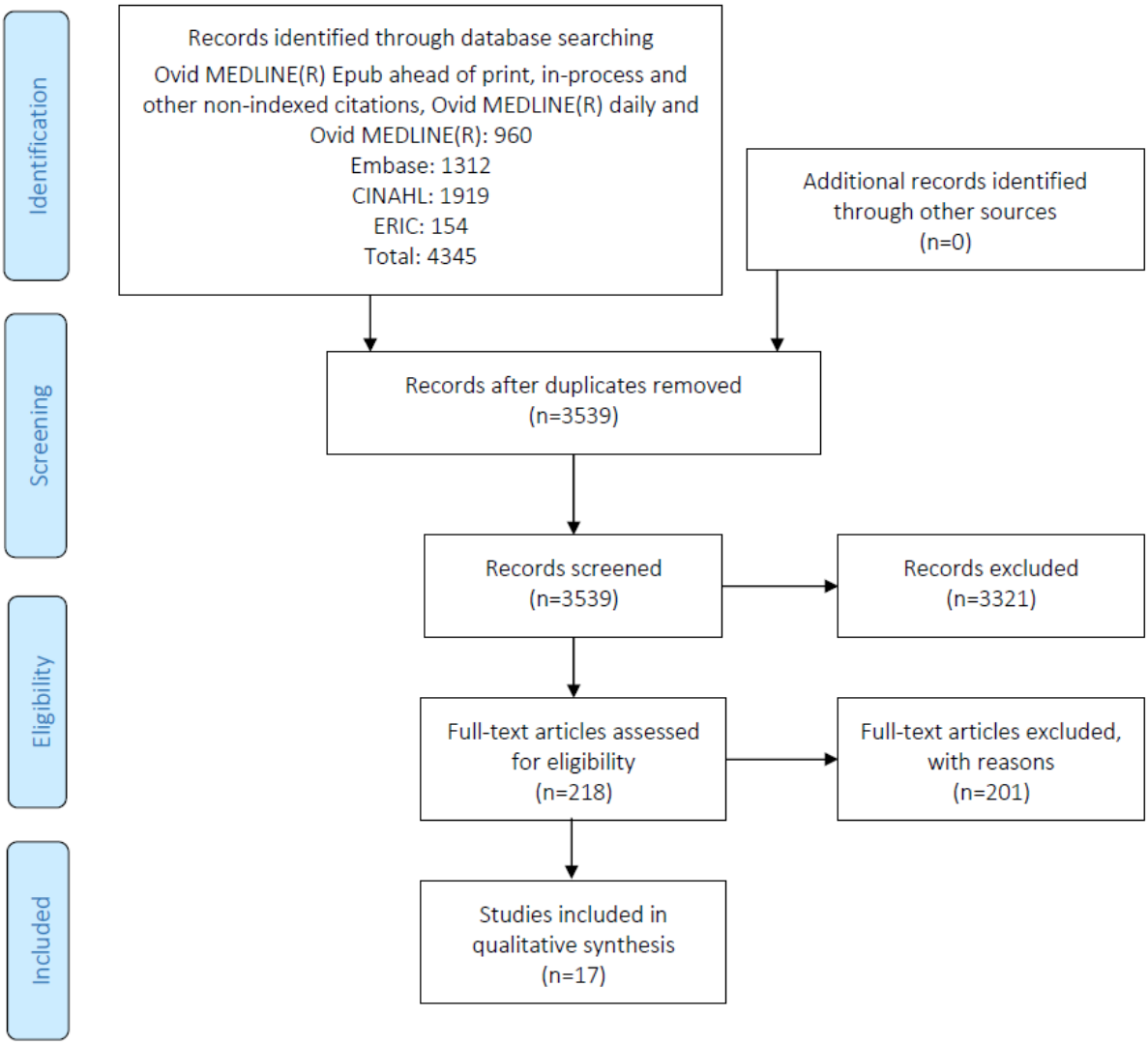
As HISs become an integral part of patient care, building an effective educational strategy may ultimately aid in the successful implementation and sustainment of an HIS. Recognizing the importance of education programs in supporting HIS implementations, this study was conducted to understand the current state of HIS education programs as reported in the academic literature. Specifically, the objective of this study is to establish a foundational understanding of the most effective strategies and approaches to enable individuals to optimally adopt and effectively use and learn from an HIS, both during and postimplementation.

Methods

Overview

A scoping review methodological framework adopting the approach by Arksey and O'Malley [12] was used to enhance the reproducibility and reliability of our findings. One of the goals of this approach was to broadly examine a topic area to map key concepts, evidence types, and current gaps in research in a well-defined field using a wide array of literature. This was an ideal starting point to better understand the landscape of research within a specific subject area. To illustrate the scoping review process, the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) diagram [13], shown in Figure 1, and the PRISMA scoping review checklist, which outlines the important milestones of a scoping review, were used [14] (Multimedia Appendix 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Review and Meta-analysis) flow diagram of scoping review results.



Stage 1: Identifying the Research Questions

To establish a baseline understanding of the most effective strategies and approaches to enable individuals to optimally use an HIS system in academic literature, this study sought to answer the following questions:

1. What education approaches led to an effective HIS implementation? What are the reported results of classroom-based, web-based, and blended learning in terms of educating staff on HISs?
2. What are the measures and outcomes used to assess the effectiveness of education and its impact on the implementation of the HIS?
3. What are the most effective approaches for enabling individuals to optimally use an HIS? How is education most effectively delivered?

Stage 2: Identifying Relevant Studies

An iterative process was used to design effective database strategies to identify eligible papers, involving several discussions with information specialists and the research team at our institution. Strategies, including subject headings, keywords, and related terms for HISs; education approaches; and training modalities were designed by a health sciences librarian for each of Ovid MEDLINE, Ovid Embase, EBSCO Cumulative Index to Nursing and Allied Health Literature, and EBSCO Education Resources Information Center and results were downloaded in November 2019 (Multimedia Appendix 2). No date or language limits were applied. HIS was defined in these searches as an electronic medical record (EMR), EHR, medical records systems, computerized medical records, electronic patient records, computer health records, computer medical records, computer hospital records, clinic information systems, electronic health medical records, and electronic hospital medical records. Relevant studies were identified through a title and abstract scan and confirmed via a full-text review. For the study selection, see the PRISMA diagram (Figure 1).

Stage 3: Study Selection

Peer-reviewed journal papers eligible for inclusion met the following criteria: (1) examined educational approaches (ie, classroom, instructor-led, web-based training, e-learning, and hybrid learning), (2) discussed HIS systems (EMR, EHR, Clinical Information system, etc), (3) discussed the effectiveness

of different approaches in educating staff to use the HIS, (4) described an educational program related to HIS, and (5) ensured education be conducted in a hospital setting. Initially, only English papers were included; however, as there were no papers found in other languages, no papers were excluded. All study designs, whether quantitative or qualitative studies and papers found in reports and research papers except for viewpoint papers, such as editorials, were included. Studies conducted in an academic setting and papers that did not meet the aforementioned criteria or describe an HIS system were excluded. Five coauthors independently reviewed the paper to determine each study's eligibility, and in cases of uncertainty, a senior reviewer with expertise in the topic of HISs was consulted.

Stage 4: Data Items and Data Collection Process

A standardized charting form was developed to capture the following domains: study details (type of study, year, and country), the objective of the study, study design (if applicable), study participants, intervention, study outcomes, and main results of the study. The PICO (Patient Problem or Population, Intervention, Comparison or Control, and Outcome) framework was used to capture details of the study, where the study outcomes were categorized using the Kirkpatrick-Barr [15] framework of educational outcomes, shown in Textbox 1. This framework was selected, as it provided a standardized method of categorizing the type of educational outcomes reported by each paper.

Textbox 1. Kirkpatrick-Barr framework of educational outcomes.

Level 1: Learners' reaction

- Learners' perspectives on the learning experience and satisfaction with the educational program [15]

Level 2a: Modification of attitudes and perceptions

- Changes in attitudes and perceptions toward patients or clients and their condition, circumstances, care, and treatment [15]

Level 2b: Acquisition of knowledge and skills

- Changes in knowledge and skills [15]

Level 3: Change in behavior

- Changes in behavior of participants' transfer of learning to their practice setting and changed professional practice [15]

Level 4a: Change in organizational practice

- Wider changes in the organizational practice and provision of care as a result of an education program [15]

Level 4b: Benefits to patients or clients

- Improvements in health or well-being of patients or clients attributable to an education program [15]

Stage 5: Synthesizing and Reporting the Results

A qualitative *narrative review* approach was adopted, and the authors independently reviewed all 17 studies. The educational outcomes reported in each paper were categorized using the Kirkpatrick-Barr framework, which also helped inform the thematic analysis. The findings were synthesized, and a thematic analysis approach was used to critically analyze the papers and develop a coding structure. Emerging themes were identified, compared, and consolidated by 3 authors. The consultation phase provided an opportunity to validate the findings and critically examine the inconsistencies or lack of clarity evident across the papers reviewed. Discussions and consultation with content experts within our team enabled us to further iterate and contextualize the themes.

Results

Search Results

The initial database search yielded 4345 papers. Once duplicates were removed, titles and abstracts of 3539 unique citations were identified. We screened these papers and identified 218 citations based on broad relevance to the topic area. The 218 abstracts then went through the second round of scrutiny against the inclusion criteria, and 33 papers were selected for full-text review. Following further inspection, 16 papers were excluded, as they did not meet the inclusion criteria. Table 1 describes the characteristics of the studies included in this study.

Table 1. Study characteristics (N=17).

Study characteristics	Value, n (%)	References
Country of publication		
United States	12 (70)	[1,4-7,16-22]
Australia	1 (6)	[23]
Namibia and Tanzania	1 (6)	[24]
United Kingdom	1 (6)	[25]
Denmark	1 (6)	[26]
The Netherlands	1 (6)	[27]
Research method		
Literature review	3 (18)	[25-27]
Questionnaire or survey	4 (23)	[6,19,21,23]
Semistructured interview	1 (6)	[9]
Mixed method	6 (35)	[4,5,16,17,20,22]
Quasiexperimental	2 (12)	[1,24]
Case report	1 (6)	[18]
Year of publication		
2016-2019	5 (29)	[5,23-26]
2010-2015	9 (53)	[1,4,6,7,16,17,19,22,27]
2006-2009	3 (18)	[18,20,21]
Education approach[a]		
In-classroom training	5 (38)	[5,6,16,18,20]
e-Learning	1 (8)	[22][b]
Blended learning	5 (38)	[1,17,19,23,24]
Classroom versus blended learning	1 (8)	[4]
Simulation training	1 (8)	[19]

[a]Literature reviews were excluded.

[b]Did not provide evaluation outcomes.

Research Question 1: Reported Results of Classroom-Based, Web-Based, and Blended Learning in Terms of Educating Staff on HISs

On the basis of the thematic analysis, this study identified 3 major themes across the different educational approaches that led to a more effective HIS implementation:

- Invest in engaging and understanding learners in the clinical context
- Maximize the transfer of learning to care
- Continuous and agile evaluation designed to meet the emerging demands of the clinical environment

In addition, the themes addressed in this study encompass fundamental elements that may be used to guide educational design and support developmental expertise in clinical environments (Table 2).

Table 2. Elements associated with the 3 major themes identified in the thematic analysis.

Theme and element	Definition of element	Studies using an element
Invest in engaging and understanding learners in the clinical context		
Assessment of individual, team, and organization capabilities	• Assess appropriate learning needs • Assess computer literacy • Ascertain clinical background and role • Provide education based on customized workflows (training based on the work environment and department needs, eg, inpatient provider, ambulatory provider)	[1,16,21,23,26,27]
Maximize the transfer of learning		
Practice and problem-based learning	• Use a problem-based approach to learning instead of task-based learning (ie, learning built on predefined tasks addressed in clinical practice). • Integrate hands-on experience to enable learner empowerment.	[1,4,6,7,17,23,25,28]
Integrate learning into practice	• Create opportunities to integrate HIS learning into practice. • Engage learners in their clinical context. • Employ simulation- and scenario-based learning techniques (real-world uncontrived experience). • Provide opportunities to create tools and items that can be used in clinical practice (eg, creating personal preference lists of frequently used orders, creating a patient-specific care plan).	[1,4,16,22,25,26]
Enhance practice improvement and performance	• Ensure the learning time is as close to the launch of the HIS as possible. • Adopt longitudinal approaches to training. • Identify and integrate super users as part of education planning. • Collaborate with and learn from clinical champions.	[6,7,17,23,25,26]
Continuous and agile evaluation designed to meet the emerging demands of the clinical environment		
Evaluation and feedback	• Being agile to meet the emerging demands of the clinical environment • Understanding how health care providers perceive HIS education and eventually its application • Adaptability and enhancement of curriculum revisions	[5,6,21]

Invest in Engaging and Understanding Learners in the Clinical Context Theme

The *Invest in engaging and understanding learners in the clinical context* theme includes both understanding learners' needs and understanding what incentives may best enable learner participation in HIS education.

Several papers have identified the importance of assessing the current capability level of learners in terms of clinical and technical skills to tailor learning accordingly, including providing adequate staff and time for education. Many health care providers received their education as health care providers before information technology became ubiquitous and, therefore, may lack the essential technical skills required to use an HIS effectively [1]. Benwell et al [23] reported that learners often felt that the HIS training sessions failed to address their learning needs, as the program was either too simplistic or advanced. This point was further reinforced in a study that examined traditional forms of education approaches, assuming that each health care provider shared the same knowledge and skill level. However, the authors discovered that the skillset of care providers differed based on their experience and educational levels [26,27]. In another peer-reviewed paper, the educational team at a hospital designed a course for health care providers with average computer skills.

> *The curriculum timing was based on an end-user with average computer skills: someone able to use a mouse and familiar with windows functionality. The design was frustrating to the advanced computer users because they were "slowed down" and frustrating to students with minimal skills because content was covered much too fast. Most staff attended training in 8-hour blocks, compounding this issue. Evaluations indicated that the number of hours spent in the classroom was over-powering and not conducive to learning. Staff reported feelings of burned out and too much content to absorb.* [21]

Previous experiences with computers and the perceptions that learners bring with them regarding the value of technology

influence their receptiveness during training sessions [21]. Edwards et al [4] stated that pretraining enabled those health care providers who needed it to gain baseline familiarity and improved their performance. Furthermore, Bredfeldt et al [1] reported:

> *Trainees who are proficient at problem list management may have already reached the functional ceiling, leaving no room for improvement. In contrast, while some training participants were very practical at medication list management, there was still room for improvement.* [1]

Adequate technology literacy and general competencies of health care staff have been identified as critical factors for implementing HIS systems [16,26]. Without these, further education sessions may be required to support health care providers before they engage in learning specific to HIS.

Four papers discussed the significance of providing incentives to help health care staff expand their information technology competencies to encompass new skills [1,6,17,18]. A study by O'Brien [18] asserted that to cover the expense of adequate training and replacement nurses, the organization provided incentives to staff to complete their training on days they were not scheduled to work. In addition, the organization has developed an incentive program to encourage staff to fill in for colleagues who were participating in training or to attend a training class in the evening or weekend at full pay. In another study, providers were recognized for their time by being eligible for continuing medical education credits [1,6].

Maximize the Transfer of Learning to Care Theme

The theme of *Maximize the transfer of learning to care* encompasses providing hands-on practice, integrating real-life case scenarios, engaging key stakeholders and staff champions, and scheduling education sessions close to the actual use of the system, all of which contribute to a learner-centric approach and a more successful HIS implementation (Table 2).

Many papers stressed the importance of providing learners with significant amounts of time to engage in hands-on activities [4]. Interaction with the HIS was a key priority among all participants across several studies [4,6,17,25]. Participants were provided with an overview of the key features of the system, using a combination of lectures and practical exercises, thus enabling the learners to gain hands-on experience using the new system [25]. In addition, hands-on activities enabled health care providers to gain more practice and become familiar with the system and have an opportunity to ask questions [6].

> *Class participants indicated that the hands-on exercises were the most useful portion of the class, and they appreciated the ability to build things in class that could be used in the clinic.* [1]

This finding corresponds with concerns from health care managers that learners should be competent in using HIS in their work setting [4]. By providing an opportunity to deliberately practice with the system, active learning is encouraged, thereby increasing learners' confidence [4]. Nicklaus et al [17] described that hands-on experience allows for the practical application of the concept and enables distraction-free instruction. This approach encourages learners to set the pace of their learning without becoming overwhelmed with a lot of content in a short period of time while increasing their confidence and competency levels [17]. By promoting a self-regulated learning principle, effective learning can be achieved by empowering learners to take part in their own learning process, set their own goals, and challenge their critical thinking skills [17]. Self-regulated learning is defined as a cyclical process that enables the learner to guide their goal-directed activities, evaluate their performance, and then reflect on the outcomes [28]. McAlearney et al [7] asserted the following:

> *...training programs that include opportunities for learners to observe others using the EHR system, and those that provide active learning opportunities, should enhance the learning process because they give learners opportunities to develop positive perceptions about their own abilities related to using the EHR.* [7]

This point was reinforced in a study by Bredfeldt et al [1], stating that classroom-based training and hands-on activities were associated with the improved utility of using the new system. The authors noted that a live EHR environment allowed staff and clinicians to build tools they could use when they returned to the clinic and the use of ancillary resources [1]. Through continuous interaction with the system, learners reduced their cognitive effort associated with performing the task [16].

In addition, several studies have found that educational programs that incorporate real case scenarios, with an emphasis on clinical workflow enhance outcomes [25,26]. Interactive scenarios presented with a mini case study highlighted the importance of new HIS elements, enabling the care providers to better understand the new updates involved with the HIS and the manner in which they needed to document [22]:

> *A hybrid teaching method that entailed both e-learning and a supplemental education session providing face-to-face personal communication, case examples, and examples of errors improved timeliness, completion and accuracy of nursing documentation significantly.* [16]

Studies have highlighted the importance of developing expertise-specific scenarios that are relevant to health care providers [25]. The use of various approaches may appeal to individual learning needs, with learners appreciative of relevant clinical scenarios in particular [25]. The scenarios were designed to reflect the daily workflow process and enabled learners to visualize how the HIS can potentially be used in their work environment [22]. Furthermore, the scenarios provided each learner with further exposure to the workflow and an opportunity to critically reflect through the different processes [22].

The literature findings described the importance of and positive changes in engaging key stakeholders and staff champions in achieving targeted organizational change. Super users (expert users who have received supplementary education and are capable of educating other staff) were found to play a critical role in providing unit-level assistance and reducing the need

for expensive external education and training [18,25]. They act as facilitators in each area or department, supporting and educating new staff [18,26]. McAlearney et al [7] reported:

> *...when learners observe others successfully using the EHR, their efficacy expectations are increased because of their corresponding beliefs that they also possess the capabilities to master the EHR system.* [7]

The authors emphasized the importance of positive behavior modeling (role modeling from peers in the work environment), which demonstrates effective approaches to help overcome these challenges. They acknowledged that engaging champions and super users may foster transformative learning and contribute to a learner-centric approach. In a study on training programs that leverage the skills of super users, it was reported that they contribute to better learning outcomes and meaningful use of the HIS [7]. Pantaleoni et al [6] described that the success of the training program was also attributed to the lead physician, as they provided guidance on the clinical context and knowledge of institutional workflows, such as the number of distinct provider workflows and how to group providers in a training session [6]. In addition, physicians were involved in the design and delivery of the training communication for hospital and staff leadership [6]. Pantaleoni et al [6] asserted that it is vital that the super user has an interest in education, institutional knowledge, and good communication skills.

In addition to the engagement of organizational stakeholders and staff champions, formalized education scheduled close to the actual use of the system was identified as beneficial to end users [23,26]. The study findings suggested that health care providers benefit from formal education only when it occurs in close proximity to their use of the HIS. Pantaleoni et al [6] emphasized the following:

> *Training classes should be offered 2 to 8 weeks prior to the change. Training that occurs greater than 8 weeks will likely not be remembered by the end-user.* [6]

It has been noted that education delivered too early or too late could potentially waste resources and raise frustrations among staff [23,26]. Moreover, the authors stated that following 30 days of unit-based experience, most staff ultimately exhibit a similar skill level. Education that is scheduled in close proximity to the time of end-user use may facilitate the greatest impact on performance and knowledge acquisition [23,26]. Furthermore, formalized (instructor-led) education may not be needed for all learners, as some participants reported formal education to be inefficient and of little value; however, daily exposure to the HIS improved their performance [23].

Continuous and Agile Evaluation Designed to Meet the Emerging Demands of the Clinical Environment Theme

A review of the studies suggested that continuous evaluation supports an agile approach to meet the emerging demands of the clinical environment (Table 2).

Pantaleoni et al [6] described the need to conduct an evaluation of an HIS educational program:

> *We then conducted an evaluation of a pilot implementation of the eLearning course to ensure that the resources matched needs; were understandable, usable, and useful; and contributed to quality improvement of future HIS eLearning resources.* [6]

Furthermore, McCain [21] stressed the value of evaluating educational programs continuously to identify course strengths and weaknesses in stimulating curriculum revisions. The authors emphasized that all HIS-related training and education programs should be continually updated to stay abreast of the evidence base and innovations.

> *Future work must include expansion and optimization of the current modules, and targeted dissemination to support uptake in appropriate settings. If evidence-based strategies for training providers Health Information Technology (HIT) are lacking, appropriate and effective use of these technologies will be limited, and many costly and potentially powerful HIT projects may fail to improve the quality of healthcare.* [21]

Hence, the evaluation allowed educators to understand the need for change in work processes and practices and an opportunity to establish mechanisms to share learning across the organization.

Importantly, although continuous evaluation can lead to learner-centric education delivery, HIS data can also be leveraged to prioritize interventions for system optimization and workflow redesign and to identify struggling learners who may require additional training or support [5]. In a study by Kadish et al [5], clinicians requested individualized training after several rounds of group training to improve their own efficiency in the EMR. Ensuring that the educational content was relevant and applicable to all learners was challenging, as individual skill sets in using the system varied among care providers. This study accentuated the importance of capturing data over time to inform continuous and personalized assistance to optimize the use of the HIS after initial training. Individualized education ensured that educators were able to adapt the content to accommodate the diversity of clinical practice at the individual and group levels to improve the competency and confidence of learners in the use of HIS to find clinical information [5].

Research Question 2: Measures and Outcomes Used to Assess the Effectiveness of Education and Its Impact on the Implementation of HIS

Twelve papers presented the results of their training evaluation [1,4-6,16-21,23,24]. As training approaches and outcomes varied across studies, each approach will be briefly discussed (Table 3), followed by measures and statements of education outcomes associated with each educational approach. The classification of educational outcomes will be guided by Kirkpatrick-Barr.

Table 3. Summary of the 12 studies that assessed the effectiveness of the education program.

Training approach and author	Measures	Outcomes
Classroom training		
Pantaleoni et al [6]	• Survey assessed providers' overall training experience, including trainer preparedness, course design, handouts, and the learner's overall readiness to use the system	• High physician satisfaction with the program *(level 1)* • Positive effect on confidence in knowledge acquired *(level 2a)*

Kadish et al [5]	• Providers were sent 2 surveys:
	  • The first survey was sent before training and used a 5-point Likert scale to measure confidence in the EMR[a] overall and in 5 key activities.
	  • Immediately after training, a second survey was sent to participants to evaluate the session and to gauge confidence in the same activities.
	• Changes in time spent in various EMR activities before and after training were compared using a paired Wilcoxon test.
		• Participants reported an increase in confidence across all activities *(level 2a)*, and almost all providers agreed that the training enhanced their efficiency (perceived; *level 3)*.
		• A reduction in the overall time in the EMR system was observed. Participants reported becoming more efficient with the use of the EMR *(level 3)*.

Evatt et al [16]	• Nurses completed a knowledge and attitude survey before and after education session:
	  • 10-item researcher-designed instrument
	  • Questions assessed knowledge regarding timeliness policy, area content, and information within areas
	  • Likert scales: testing attitude toward the completion of the EHR[b] nursing admission assessment
	• Sampled charts of patients admitted to the 2 units before and after completion of the sessions:
	  • To evaluate the timeliness of completion, the time (in minutes) from patient admission to the unit to submission of the nursing admission assessment to the EHR was determined.
	• Accuracy of documentation was assessed with regard to the accuracy of the past medical history.
		• Nurses' attitudes *(level 2a)* and knowledge *(level 2b)* regarding completion of the EHR nursing assessment admission assessment improved significantly.
		• Following the educational session, the mean time to completion of the EHR nursing admission assessment decreased *(level 3)*.
		• Timeliness, completeness, and accuracy of assessment documentation (after the session) were improved significantly after use of a hybrid approach *(level 3)*.

O'Brien [18]	• Physician survey was conducted 2 months after the physician order-entry go-live date.	• 90% agreed that the EMR system made it easier for them to do their work *(level 1)*. • Medication errors caused by illegibility and transcription were eliminated completely *(level 3)*. • Staff also have found that the EMR system makes their jobs more efficient *(level 3)*. • Patient satisfaction scores for the overall satisfaction with care climbed to their highest levels *(level 4b)*.
Kraus et al [20]	• Measure the level of adoption of CPOM[c]: percentage of the utilization of HIS[d] by CPOM of physicians.	• Physician adoption reached the first-year goal of 40% physician entry in the first month and stabilized at 75% within a year. • The impact has been noted in pharmacy, where the average time from order to pharmacist verification decreased from 90 min before CPOM to 17.9 min a year later *(level 3)*. • The resulting order sets and their increasing use in clinical care can be considered another measure of success.
Blended learning		

Training approach and author	Measures	Outcomes

Bredfeldt et al [1]	• Evaluated 2 outcome measures in the EHR data for 6 months before and after training:
	  • Proportion of visits in which either the problem list or the medication list was modified
	  • Modifying the problem list included adding or deleting problems from the problem list or attaching comments to existing problems on the list
	• Modifying the medication list included marking medications as chronic, removing inactive medications, or marking the medication list as reviewed
		• Training was related to a small but significant increase in the use of key EHR capabilities:
		  • Participants increased their use of both the problem list from 22% to 24% of visits and their use of the medication list from 41% to 45% of visits after the education session *(level 3)*.

Benwell et al [23]	• Questionnaire (10-point Likert scale and dichotomous scale) • To rate their self-perceived skill level using computers in general versus BOSSnet (digital medical record), the usefulness of the ICT[e] training session and their willingness to train others • The time taken to complete all tasks (efficiency) and the number of incorrect mouse clicks (accuracy) used to complete each task were recorded during the education session.	• A significant improvement in both efficiency and accuracy for all participants during the session was observed *(level 2b)*. • The greatest improvements in task performance followed daily ward–based experience using BOSSnet rather than formalized training.
Nicklaus et al [17]	• Observations to measure the effectiveness of training	• Learners were satisfied with the learning laboratory as it provided an opportunity for them to practice and understand the system *(level 1)*. • They reported that scenario-based practice time stimulated realistic documentation and that they began to better comprehend the information.
Rudd et al [24]	• The primary evaluation outcome was knowledge gain resulting from the completion of the blended e-learning course, measured by differences in posttest and pretest scores. • Secondary outcomes included achievement of a 70% passing score and participant satisfaction with e-learning module content, format, and delivery.	• Respondents reported high satisfaction with the overall content of the course and with the e-learning modules *(level 1)*. • Blended e-learning course participants gave positive feedback about the course structure *(level 1)*, and their knowledge of HIS competencies. • Participants experienced strong learning gains in both, although learning gains were somewhat greater in the in-person course (level 2b).
McCain [21]	• Likert scale and open-ended questions • Rating how well objectives were met and clarity of information, in addition to sharing course strengths and suggestions for improvement	• Reduction in training time was observed. • Participants liked being able to complete the training at their own pace and immediately practice the information learned *(level 1)*.
Classroom training versus BL[f]		

Training approach and author	Measures	Outcomes

Edwards et al [4]	• Satisfaction with HIT[g] training was assessed using a pre-existing, web-based, anonymous self-report survey.
	  • The instrument consisted of 13 questions: 9 questions focused on satisfaction with course execution, instructor quality, and usefulness of materials
	  • Ratings from these 9 questions were summed to form a single satisfaction score.
	  • Three subjective questions: identify the most and least valuable information and materials experienced during training
	  • Additional question elicited general comments and suggestions
		• Learners were equally satisfied with both methods: instructor-led and blended learning BL *(level 1).*
		  • Instructor-led participants found the training to be valuable *(level 1)*, particularly the system functions and navigation.

Simulation training		
Vuk et al [19]	• Questionnaires on a 7-point Likert scale before and immediately after simulation training • Assessed their perceptions about the importance of EMRs in improving patients' safety and their confidence and preparedness level to use EMRs	• Simulation training enhanced physicians' and nurses' levels of self-confidence and preparedness to use EMRs *(level 2a).*

[a]EMR: electronic medical record.

[b]EHR: electronic health record.

[c]CPOM: computerized physician order management.

[d]HIS: health information system.

[e]ICT: information computer technology.

[f]BL: blended learning.

[g]HIT: health information technology.

Discussion

Current State of HIS Education Programs

This study identified critical knowledge gaps in our understanding of the most effective strategies and approaches in enabling health care providers to optimally use an HIS system. We only identified 17 studies that examined the effects of different education tactics in a hospital environment. This paucity of literature indicates the maturity of the topic area and emphasizes the need to establish a baseline understanding of HIS education strategies during and post implementation. This study provides an understanding of the current landscape of these programs and important insights into successful education development and improvement.

HIS education and training have been identified as potential key facilitators in ensuring effective and optimal use of technology and can have a positive impact on HIS implementation, efficiency, and patient care. Competency in HIS is now an essential clinical skill, and health care providers and staff who lack proficiency and efficiency may face challenges in performing clinical tasks [29]. Despite this, many organizations underestimate education and training needs and the time required for effective education [10]. Unsuccessful transitions are also because of a lack of understanding of what learners expect to gain from training and failure to link training to HIS implementation. Furthermore, a lack of adequate education may increase the risk of users creating workarounds that limit the advantages of HIS and potentially hold the organization back [10].

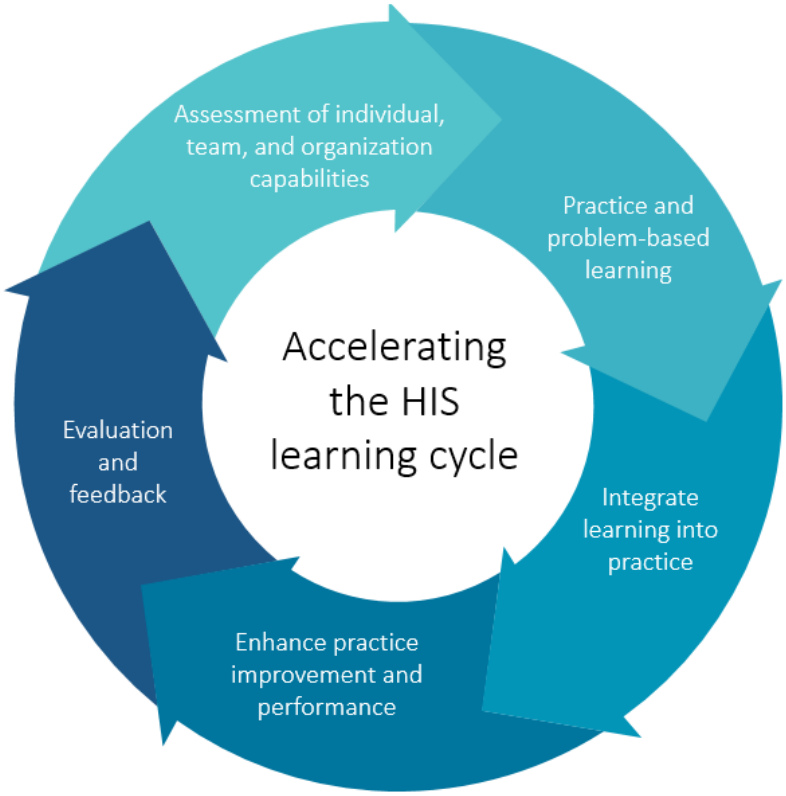
Upon review of the numerous instruments employed in the reviewed literature, it is apparent that no standardized tools have been adopted yet to assess learner outcomes of education programs. The majority of authors used self-constructed, nonvalidated scales and defined their results in qualitative terms. This limits all future efforts to compare and analyze evidence on the effectiveness of the HIS curriculum. More studies with standardized outcome measures and assessment tools are required to support recommendations on the most effective approaches in enabling providers to optimally use an HIS.

Fortunately, HIS training is beginning to embrace a learner-centric paradigm, and HIS education can be informed by existing educational frameworks such as Kirkpatrick and Moore. For example, Kirkpatrick and Moore's frameworks are focused on evaluating health care provider education and have become a commonly cited reference when assessing educational outcomes. Although these frameworks provide a good source of reference for evaluating the impact of learning and development, these frameworks do not provide sufficient direction for designing and sustaining an HIS education program. In particular, the frameworks do not take into account a wide array of factors, including those associated with the organization, individuals, teams, and the overall design of the

education itself, all of which can influence the effectiveness of the educational program before, during, or after education delivery. Similarly, the experiential learning theory by Kolb focuses on learning at the individual level, a helpful but insufficient perspective to guide planning for HIS education across an organization. Finally, and importantly, none of these frameworks are grounded in evidence unique to HIS education.

Given many of the considerations that are unique to HISs and the limited applicability of existing educational frameworks, the authors developed a new framework for HIS education that guides the adoption of the most effective education strategies used to equip health care providers with the skills required to work effectively in a clinical environment. To address these gaps, the Accelerating the HIS Learning Cycle Framework in Figure 2, with the 5 fundamental elements, was developed from the data from this study.

Figure 2. Accelerating the health information system learning cycle framework. HIS: Health Information System.



Research Question 3: Most Effective Approaches in Enabling Individuals to Optimally Use an HIS

Assessment of Individual, Team, and Organization Capabilities

The assessment of the current capability of learners (individuals and teams) before the delivery of training appears to be a critical facet of effective HIS implementation. As learners progress in their environment and acquire new skills and methods of engaging in their ecosystem, education should be adapted to address their learning needs. Hence, it is worth acknowledging that there may be different levels of HIS knowledge and adoption among providers. Terry et al [30] and Harton et al [31] noted that although there is variability regarding the influence of previous technical knowledge on perceptions of EMR adoption, learners with lower digital literacy may require extra sessions to learn the functionality of the computer and the processes necessary for clinical practice. This finding is particularly important as health care systems evolve to leverage HISs to support clinicians' adaptation to the new workflows and integration of this technology into their clinical practice. Despite the value placed on interprofessional learning, the literature has primarily focused on individual learning and a dearth of evidence has been provided on team-based learning. Future research exploring interprofessional education could provide greater insights into designing an effective educational program.

Culture pervades learning, and to meet the needs of diverse learners, issues revolving around the social and cultural dimensions of task design, structuring of content, and communication channels must be considered when designing a curriculum [32]. Culturally responsive pedagogy (CRP) recognizes learners' differences and stresses that the cultural congruence of an instructional environment increases learners' success [33]. A study from the University of British Columbia revealed that learners who are culturally diverse have a tenuous relationship with institutions that focus their curriculum on traditional, Eurocentric, and normative approaches [33]. These approaches tend to neglect learners from marginalized cultural backgrounds by disregarding their cultural habitus, leading to a cultural discontinuity for learners and the organization [33]. Rijal [34] reported that learning organizations moving toward a culture that encourages openness, creativity, experimentation, and tolerance for mistakes will enhance learning outcomes. An

important part of a learning organization is being able to create new knowledge and use it to capitalize on new opportunities [35]. CRP has been hypothesized to connect all facets of learning with each other on emotional, social, mental, and physical levels [33]. In this framework, CRP is an important element in understanding individuals' and organizations' needs.

Practice and Problem-Based Learning

This study surfaced the relevance of problem-based learning in HIS education and training. A majority of the papers focused on a task-based learning approach to training, where learning is built based on predefined tasks addressed in clinical practice. Unfortunately, a task-based approach can lead to learners being dependent on instructors and support tools to provide them with guidance to perform the task, rather than encouraging them to critically think about how to approach the task [36]. Focusing on a narrow learning parameter without examining the larger context of the HIS system may not be sufficient to successfully transform learners from the existing approach to the new electronic documentation [17]. Problem-based learning provides a promising avenue for delivering an optimal learning experience while fostering an active independent learning attitude in learners. This learning approach enables learners to internalize knowledge through a process of solving clinical problems and stimulate deeper thinking with an emphasis on *how and why* questions [36]. Education that is at the appropriate skill level of learners and focused on a problem-based learning approach may encourage learners to critically reflect and attempt to understand not only the tasks themselves but also the concepts and mechanisms underlying the tasks.

This study underscores the importance of incorporating hands-on practice as part of education to increase learners' confidence and competency in successfully using the HIS system. In a review by Younge et al [37], the authors asserted that hands-on practice addressed the learner's level of computer literacy, which also relates to their ability regarding the ease or difficulty of using HIS. Similarly, Youssef [10] contended that hands-on experience enables learners to develop realistic expectations of what the HIS is able to offer. Thus, learners are able to strengthen the connection between personal experiences, learning content, knowledge, and a concrete task, resulting in better comprehension of abstract concepts [38].

Integrate Learning Into Practice

Iterative assessment of learners' performance with new scenarios enables learners to demonstrate their knowledge and their competency in using their knowledge to deal with the new and more difficult cases being presented. Practice-based learning strengthens learners' knowledge integration and application in a real-life setting [38]. Younge et al [37] noted that educating with materials, which provide opportunities for active learning and using assessments that evaluate what learners know (efficiency) and how they use existing knowledge to solve new problems helps to foster adaptive expertise. Learners highlighted real-life scenarios as a way to augment critical thinking by engaging in discussions. Hence, digital learning resources must be designed in a manner that offers better immersion while not increasing cognitive load [38].

Use of hands-on learning is consistent with the main phases of the experiential learning theory by Kolb [39]. The incorporation of hands-on practice and real case scenarios provides learners with an opportunity to deal with the workflows in clinical practice, which occurs when clinicians engage in an uncertain and unfamiliar context and allows learners to take an active role in the learning process. Using hands-on practice and case studies built on the 4 stages of the experiential learning theory would elicit evidence for changes in the cognitive process, learning, and behavior. This is critical in underpinning the design of an HIS curriculum and the role of educators and learners.

The clinical environment is an ideal setting to identify knowledge and skill gaps and then pursue learning with colleagues and instructors in a venture to fill these gaps. The authors advocate for the development of an interprofessional community of practice (CoP) as it facilitates the sharing of best practices and allows for the creation of new knowledge to advance the domain of HIS in clinical practice [40]. A CoP is built based on the assumptions of co-participation, where all learners from varied geographical locations engage in the activities of the community intending to facilitate meaningful learning [41]. As long as learners and instructors are present and engaged, an online community will evolve dynamically to meet their specific needs [41]. The digital space allows participants to share their experiences and knowledge in creative ways to cultivate new approaches to problems. When designing the HIS education strategy, this element of sustainability should be considered to enable individuals to optimally adopt and effectively use an HIS.

Enhance Practice Improvement and Performance

Another critical element as part of an educational strategy is identifying and engaging super users early in the project to help foster learning and understand the value of the HIS in clinical settings. Engaging super users in the development of an educational program enables the content to be tailored to specific provider needs, which, in turn, will contribute to the overall HIS adoption and successful implementation. The identification and engagement of champions and super users are rooted in the diffusion of innovation theory developed by Rogers [42]. Super users and champions are early adopters and innovators who adopt the HIS very early and take part in the dissemination of the new idea within the organization. They use the communication channels established to influence people's attitudes and accelerate the rate of adoption [42].

Evaluation and Feedback

Evaluation plans should not only evaluate the efficacy of the initial training but also use this information to inform plans to address the ongoing learning needs during and post-HIS education and implementation. In addition, ongoing evaluation can provide insights into emerging learning needs not only about the HIS system but also about emerging practice gaps and variations, which can help refine the goals and objectives and guide the implementation of the most effective education strategies. Through evaluation associated with HIS education and implementation, areas of significant strength can also be identified; for example, one part of the organization may demonstrate exceptional use of the HIS in practice. Positive

deviance allows the organization to identify the top performers and foster analysis and discussion of such performance to elevate performance among other groups within the organization [43]. This approach characterizes not only the processes and practices that exist in top-performing groups but also the context in which they are implemented, such as the organizational culture and norms of behaviors [43]. Evaluation can be used to maintain and garner support for an education program in addition to assessing learner achievement.

Limitations

The findings of our scoping review should be examined in the context of the following limitations. Due to the nature of the scoping review, the quality of each study was not assessed. The age of the literature and the gap in publication dates may curtail the validity of the findings concerning the current landscape, as many of the previous papers were published in a different health care climate. Moreover, based on 5 studies that assessed the third level of Kirkpatrick, one was a perceived outcome in behavior, and only one of the studies assessed the highest level of Kirkpatrick-Barr (results). Given the nature of the topic being investigated, we excluded studies that discussed HIS education in academic institutions (eg, universities). Another limitation of this study is that we cannot confirm that we did not miss any relevant studies as the literature on the most effective approaches in enabling individuals to optimally use an HIS is heterogeneous. There is no standardized terminology in educational research, and the term used to describe the same ideas may vary depending on the author, thus limiting the retrieval of papers comprising important findings. The acceleration of the HIS learning cycle framework is emergent from this study and itself has not yet been validated.

Conclusions

This study supports the development of an HIS learning framework that educators can use to guide the design and development of HIS education and training during and postimplementation. Given the advances in HISs, health care organizations should be equipped with the essential tools to deal with the turbulence that embodies digital ecosystems and research into all facets of education that prepare health providers, teams, and the organization as a whole, for the rapidly changing nature of clinical environments. This framework is a novel addition to the literature and needs to be pilot tested to evaluate their feasibility and efficacy in health care education. We posit that to successfully transform care providers to use the new technology, best practices and training principles in HIS education that harness the nature of transformative learning must be pursued. Future efforts should examine the effectiveness of interprofessional education interventions, as the literature predominantly focuses on individualized learning. Finally, we encourage future studies to focus on iterative learning to better understand how providers continue to learn from the HIS postimplementation about key practice gaps.

Acknowledgments

The authors wish to thank Sharon Ambata-Villanueva for her support and thoughtful feedback on previous iterations of this manuscript.

Authors' Contributions

DW led the conceptualization and design of the review, data analyses, and revised all drafts of this manuscript for important intellectual content and clarity. MA developed the search strategy and conducted the search and provided feedback on the manuscript. KF, MA, SM, TJ, and TT contributed to the identification of papers and screening. ML, TJ, and SM collaborated on the thematic analysis of the collected data, drafting, and finalization of the manuscript. DW, ML, and MT have contributed to the development of ideas that were instrumental in surfacing and maturing many of the concepts contained in this study. They also served as content experts in validating the findings and revising all drafts. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Review and Meta-analysis) scoping review checklist.
[DOCX File , 31 KB - jmir_v23i2e24691_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 38 KB - jmir_v23i2e24691_app2.docx]

References

1. Bredfeldt CE, Awad EB, Joseph K, Snyder MH. Training providers: beyond the basics of electronic health records. BMC Health Serv Res 2013 Dec 2;13:503 [FREE Full text] [doi: 10.1186/1472-6963-13-503] [Medline: 24295150]

2. AbouZahr C, Boerma T. Health information systems: the foundations of public health. Bull World Health Organ 2005 Aug;83(8):578-583 [FREE Full text] [Medline: 16184276]
3. Keshavjee K, Bosomworth J, Copen J, Lai J, Kucukyazici B, Lilani R, et al. Best practices in EMR implementation: a systematic review. AMIA Annu Symp Proc 2006:982 [FREE Full text] [Medline: 17238601]
4. Edwards G, Kitzmiller RR, Breckenridge-Sproat S. Innovative health information technology training: exploring blended learning. Comput Inform Nurs 2012 Feb;30(2):104-109. [doi: 10.1097/NCN.0b013e31822f7f7a] [Medline: 21915046]
5. Kadish SS, Mayer EL, Jackman DM, Pomerantz M, Brady L, Dimitriadis A, et al. Implementation to optimization: a tailored, data-driven approach to improve provider efficiency and confidence in use of the electronic medical record. J Oncol Pract 2018 Jul;14(7):e421-e428. [doi: 10.1200/JOP.18.00093] [Medline: 29939808]
6. Pantaleoni JL, Stevens LA, Mailes ES, Goad BA, Longhurst CA. Successful physician training program for large scale EMR implementation. Appl Clin Inform 2015;6(1):80-95 [FREE Full text] [doi: 10.4338/ACI-2014-09-CR-0076] [Medline: 25848415]
7. McAlearney AS, Robbins J, Kowalczyk N, Chisolm DJ, Song PH. The role of cognitive and learning theories in supporting successful EHR system implementation training: a qualitative study. Med Care Res Rev 2012 Jun;69(3):294-315. [doi: 10.1177/1077558711436348] [Medline: 22451617]
8. Mclean E, Sedera D, Tan F. Reconceptualizing System Use for Contemporary Information Systems. In: 15th Pacific Asia Conference on Information Systems. 2011 Jul Presented at: CIS'11; 2011 July 7-11; Brisbane, Australia.
9. Shachak A, Kuziemsky C, Petersen C. Beyond TAM and UTAUT: Future directions for HIT implementation research. J Biomed Inform 2019 Dec;100:103315. [doi: 10.1016/j.jbi.2019.103315]
10. Youssef W. Evaluation of EHR training as a catalyst to achieve clinician satisfaction with technology in acute care setting. MSc Thesis. University of Victoria. 2013. URL: http://dspace.library.uvic.ca/bitstream/handle/1828/4827/Youssef_Walid_MSc_2013.pdf?sequence=1&isAllowed=y [accessed 2021-02-02]
11. Kumar S, Aldrich K. Overcoming barriers to electronic medical record (EMR) implementation in the US healthcare system: A comparative study. Health Informatics J 2010 Dec;16(4):306-318 [FREE Full text] [doi: 10.1177/1460458210380523] [Medline: 21216809]
12. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2007 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
13. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
14. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
15. Shen N, Yufe S, Saadatfard O, Sockalingam S, Wiljer D. Rebooting kirkpatrick: integrating information system theory into the evaluation of web-based continuing professional development interventions for interprofessional education. J Contin Educ Health Prof 2017;37(2):137-146. [doi: 10.1097/CEH.0000000000000154] [Medline: 28562503]
16. Evatt M, Ren D, Tuite P, Reynolds C, Hravnak M. Development and implementation of an educational support process for electronic nursing admission assessment documentation. Medsurg Nurs 2014;23(2):89-95, 100. [Medline: 24933785]
17. Nicklaus J, Kusser J, Zessin J, Amaya M. Transforming education for electronic health record implementation. J Contin Educ Nurs 2015 Aug;46(8):359-363. [doi: 10.3928/00220124-20150721-02] [Medline: 26247658]
18. O'Brien MS. Implementation of the EPIC electronic medical record/physician order-entry system. J Healthc Manag 2006;51(5):338-343. [Medline: 17039692]
19. Vuk J, Anders ME, Mercado CC, Kennedy RL, Casella J, Steelman SC. Impact of simulation training on self-efficacy of outpatient health care providers to use electronic health records. Int J Med Inform 2015 Jun;84(6):423-429. [doi: 10.1016/j.ijmedinf.2015.02.003] [Medline: 25746460]
20. Kraus S, Barber TR, Briggs B, Kall G, Vernon TR, Moran B, et al. Implementing computerized physician order management at a community hospital. Jt Comm J Qual Patient Saf 2008 Feb;34(2):74-84. [doi: 10.1016/s1553-7250(08)34010-0] [Medline: 18351192]
21. McCain CL. The right mix to support electronic medical record training: classroom computer-based training and blended learning. J Nurses Staff Dev 2008;24(4):151-154. [doi: 10.1097/01.NND.0000320673.65824.db] [Medline: 18685473]
22. Topaz M, Rao A, Masterson Creber R, Bowles KH. Educating clinicians on new elements incorporated into the electronic health record: theories, evidence, and one educational project. Comput Inform Nurs 2013 Aug;31(8):375-9; quiz 380 [FREE Full text] [doi: 10.1097/NXN.0b013e318295e5a5] [Medline: 23774448]
23. Benwell N, Hird K, Thomas N, Furness E, Fear M, Sweetman G. Effectiveness and efficiency of training in digital healthcare packages: training doctors to use digital medical record keeping software. Aust Health Rev 2017 Oct;41(5):479-484. [doi: 10.1071/AH16090] [Medline: 27591794]
24. Rudd KE, Puttkammer N, Antilla J, Richards J, Heffron M, Tolentino H, et al. Building workforce capacity for effective use of health information systems: evaluation of a blended eLearning course in Namibia and Tanzania. Int J Med Inform 2019 Nov;131:103945. [doi: 10.1016/j.ijmedinf.2019.08.005] [Medline: 31561193]

25. Brown CL, Reygate K, Slee A, Coleman JJ, Pontefract SK, Bates DW, et al. A literature review of the training offered to qualified prescribers to use electronic prescribing systems: why is it so important? Int J Pharm Pract 2017 Jun;25(3):195-202. [doi: 10.1111/ijpp.12296] [Medline: 27488258]
26. Bygholm A. Staff training on the use of health information systems: what do we know? Stud Health Technol Inform 2018;247:191-195. [Medline: 29677949]
27. Goveia J, Van Stiphout F, Cheung Z, Kamta B, Keijsers C, Valk G, et al. Educational interventions to improve the meaningful use of Electronic Health Records: a review of the literature: BEME Guide No. 29. Med Teach 2013 Nov;35(11):e1551-e1560. [doi: 10.3109/0142159X.2013.806984] [Medline: 23848402]
28. Sanders MR, Turner KM, Metzler CW. Applying self-regulation principles in the delivery of parenting interventions. Clin Child Fam Psychol Rev 2019 Mar;22(1):24-42. [doi: 10.1007/s10567-019-00287-z] [Medline: 30788658]
29. Stroup K, Sanders B, Bernstein B, Scherzer L, Pachter LM. A new EHR training curriculum and assessment for pediatric residents. Appl Clin Inform 2017 Oct;8(4):994-1002 [FREE Full text] [doi: 10.4338/ACI-2017-06-RA-0091] [Medline: 29241239]
30. Terry AL, Giles G, Brown JB, Thind A, Stewart M. Adoption of electronic medical records in family practice: the providers' perspective. Fam Med 2009;41(7):508-512 [FREE Full text] [Medline: 19582637]
31. Harton BB, Borrelli L, Knupp A, Rogers N, West VR. Integrating traditional nursing service orientation content with electronic medical record orientation. J Nurses Staff Dev 2009;25(5):229-235. [doi: 10.1097/NND.0b013e3181ba3bb4] [Medline: 19820534]
32. McLoughlin C. Inclusivity and alignment: principles of pedagogy, task and assessment design for effective cross‐cultural online learning. Distance Educ 2006 Nov;22(1):7-29. [doi: 10.1080/0158791010220102]
33. Ragoonaden K, Mueller L. Culturally Responsive Pedagogy: Indigenizing Curriculum. Can J Higher Educ Rev 2017 Aug 27;47(2):22-46. [doi: 10.47678/cjhe.v47i2.187963]
34. Rijal S. The influence of transformational leadership and organizational culture on learning organization: A comparative analysis of the it sector. J Admin Bus Stud 2016 Jun 21;2(3):121-130. [doi: 10.20474/jabs-2.3.3]
35. Goh S. Toward a learning organization: the strategic building blocks. SAMJ 1998;63(2):15-22.
36. Takahashi Y. Problem-based learning and task-based learning: a practical synthesis. Kaohsiung J Med Sci 2008 Mar;24(3 Suppl):S31-S33 [FREE Full text] [doi: 10.1016/s1607-551x(08)70091-3] [Medline: 18364284]
37. Younge V, Borycki E, Kushniruk A. On-the-job training of health professionals for electronic health record and electronic medical record use: a scoping review. Knowledge Manag E-Learning 2015 Sep;7(3):436-469. [doi: 10.34105/j.kmel.2015.07.029]
38. Chen J. Developing a cycle-mode POED model and using scientific inquiry for a practice activity to improve students' learning motivation, learning performance, and hands-on ability. Interact Learn Environ 2020 Jan 20:1-13. [doi: 10.1080/10494820.2020.1716023]
39. Fewster-Thuente L, Batteson T. Kolb's experiential learning theory as a theoretical underpinning for interprofessional education. J Allied Health 2018;47(1):3-8. [doi: 10.1007/978-1-4419-1428-6_227] [Medline: 29504014]
40. Li LC, Grimshaw JM, Nielsen C, Judd M, Coyte PC, Graham ID. Evolution of Wenger's concept of community of practice. Implement Sci 2009 Mar 1;4:11 [FREE Full text] [doi: 10.1186/1748-5908-4-11] [Medline: 19250556]
41. Pringle R. Developing a community of learners: potentials and possibilities in web mediated discourse. Contemp Issues Technol Teach Educ Online serial 2002;2(2):218-233.
42. Greenhalgh T, Stramer K, Bratan T, Byrne E, Mohammad Y, Russell J. Introduction of shared electronic records: multi-site case study using diffusion of innovation theory. Br Med J 2008 Oct 23;337:a1786 [FREE Full text] [doi: 10.1136/bmj.a1786] [Medline: 18948344]
43. Bradley EH, Curry LA, Ramanadhan S, Rowe L, Nembhard IM, Krumholz HM. Research in action: using positive deviance to improve quality of health care. Implement Sci 2009 May 8;4:25 [FREE Full text] [doi: 10.1186/1748-5908-4-25] [Medline: 19426507]

Abbreviations

CRP: culturally responsive pedagogy
EHR: electronic health record
EMR: electronic medical record
HIS: Health Information System
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis

Edited by G Eysenbach; submitted 30.09.20; peer-reviewed by A Shachak, A Bygholm, S Nagavally; comments to author 04.12.20; revised version received 05.01.21; accepted 16.01.21; published 24.02.21.

Please cite as:
Jeyakumar T, McClure S, Lowe M, Hodges B, Fur K, Javier-Brozo M, Tassone M, Anderson M, Tripp T, Wiljer D
An Education Framework for Effective Implementation of a Health Information System: Scoping Review
J Med Internet Res 2021;23(2):e24691
URL: https://www.jmir.org/2021/2/e24691
doi:10.2196/24691
PMID:33625370

Viewpoint

One Digital Health: A Unified Framework for Future Health Ecosystems

Arriel Benis[1,2], PhD; Oscar Tamburis[3], PhD; Catherine Chronaki[4], MSc; Anne Moen[5], RN, PhD

[1]Faculty of Technology Management, Holon Institute of Technology, Holon, Israel

[2]Faculty of Digital Medical Technologies, Holon Institute of Technology, Holon, Israel

[3]Department of Veterinary Medicine and Animal Productions, University of Naples Federico II, Naples, Italy

[4]HL7 Europe, Brussels, Belgium

[5]Faculty of Medicine, Institute for Health and Society, University of Oslo, Oslo, Norway

Corresponding Author:
Arriel Benis, PhD
Faculty of Technology Management
Holon Institute of Technology
Golomb St 52
PoB 305
Holon, 5810201
Israel
Phone: 972 035026892
Email: arrielb@hit.ac.il

Abstract

One Digital Health is a proposed unified structure. The conceptual framework of the One Digital Health Steering Wheel is built around two keys (ie, One Health and digital health), three perspectives (ie, individual health and well-being, population and society, and ecosystem), and five dimensions (ie, citizens' engagement, education, environment, human and veterinary health care, and Healthcare Industry 4.0). One Digital Health aims to digitally transform future health ecosystems, by implementing a systemic health and life sciences approach that takes into account broad digital technology perspectives on human health, animal health, and the management of the surrounding environment. This approach allows for the examination of how future generations of health informaticians can address the intrinsic complexity of novel health and care scenarios in digitally transformed health ecosystems. In the emerging hybrid landscape, citizens and their health data have been called to play a central role in the management of individual-level and population-level perspective data. The main challenges of One Digital Health include facilitating and improving interactions between One Health and digital health communities, to allow for efficient interactions and the delivery of near–real-time, data-driven contributions in systems medicine and systems ecology. However, digital health literacy; the capacity to understand and engage in health prevention activities; self-management; and collaboration in the prevention, control, and alleviation of potential problems are necessary in systemic, ecosystem-driven public health and data science research. Therefore, people in a healthy One Digital Health ecosystem must use an active and forceful approach to prevent and manage health crises and disasters, such as the COVID-19 pandemic.

(J Med Internet Res 2021;23(2):e22189) doi:10.2196/22189

KEYWORDS

One Health; digital health; eHealth; medicine; veterinary medicine; environmental monitoring; education; patient engagement; citizen science; health care industry; population health management; data science; COVID-19

Introduction and Background

All science fields are strongly interconnected to each other, and each health specialty depends on others. Yet academics, researchers, and practitioners are often confined to their own knowledge, language, and expertise silos. This issue is paradoxically enhanced by the existing abundance of knowledge; large amounts of big data and the cloud age have supported short-term research in numerous domains [1-3].

One Health is an umbrella concept that encompasses all disciplines that broadly deal with human health, animal health, and the surrounding environment. The modern origins of One Health date back to 2004. One Health was introduced as part of the 12 Manhattan Principles, which called for an international,

interdisciplinary approach for preventing diseases [4], specifically animal-human transmissible and communicable diseases. Therefore, systematic perspectives on life sciences and the environment (ie, interactions and coinfluences within the environment) were brought together to design and implement programs, policies, and regulations for achieving better public health outcomes. The World Health Organization (WHO) has associated One Health with sustainable development goals [5,6]. One Health involves the evaluation and monitoring of the impact of environmental hazards on health care systems, public health, biodiversity, and food security. In summary, One Health encompasses the interconnections between humans, animals, plants, and the global ecological environment [7-10]. Therefore, One Health research requires close collaboration among health practitioners, public health workers, biologists, and environmental science specialists. One of the main aims of One Health research is to understand how potential vector-borne diseases (eg, a highly pathogenic avian influenza) spread and how they can be controlled [11,12].

The evolution of digital technology has resulted in evidence-based, accessible, upper-level, and holistic methods that are capable of accelerating biomedical research and enhancing public health efficacy. Moreover, this transformation has rapidly enhanced scientific knowledge development and improved health education, personalized clinical care, and citizen science. This has led to the growth and advancement of scientific knowledge, as well as overlaps among formal, natural, and social sciences; engineering; health; medicine; and well-being. There have been calls for integrating digital technologies into the living- and society-related aspects of various fields to enhance population health, efficiency, precision, and personalization in health care delivery. Thanks to active citizens' engagement and participation, considerable volumes of data have been generated by a multitude of systems and processed with semi–real-time data mining techniques (ie, techniques that health care data specialists have improved over time). Reflecting upon multiple perspectives during the ongoing COVID-19 pandemic can provide opportunities for revisiting and understanding the limitations of public health, human and veterinary health care, and environmental management systems [13,14].

To build upon these advances in digital technology, we introduce One Digital Health (ODH), which is a novel framework that integrates perspectives from health, care, and well-being research; health informatics research; and broad computer and data science research. ODH is based on (1) health informatics and the broadening of digital health [15]; (2) One Health [4], which provides a holistic and systemic view of health and life sciences that are prominent in human health, animal health, and ecosystem research; and (3) environmental research.

In the following sections of this viewpoint paper, we will elaborate on the ODH conceptual framework, highlight the core elements of ODH, and address specific challenges for the future. Examples of COVID-19–related aspects and how these aspects relate to the ODH framework will also be considered and discussed. We will also discuss recommendations for raising awareness, capacity building in ODH, and the proposed multidisciplinary European Federation of Medical Informatics (EFMI) ODH working group, to promote synergy across One Health scientific fields.

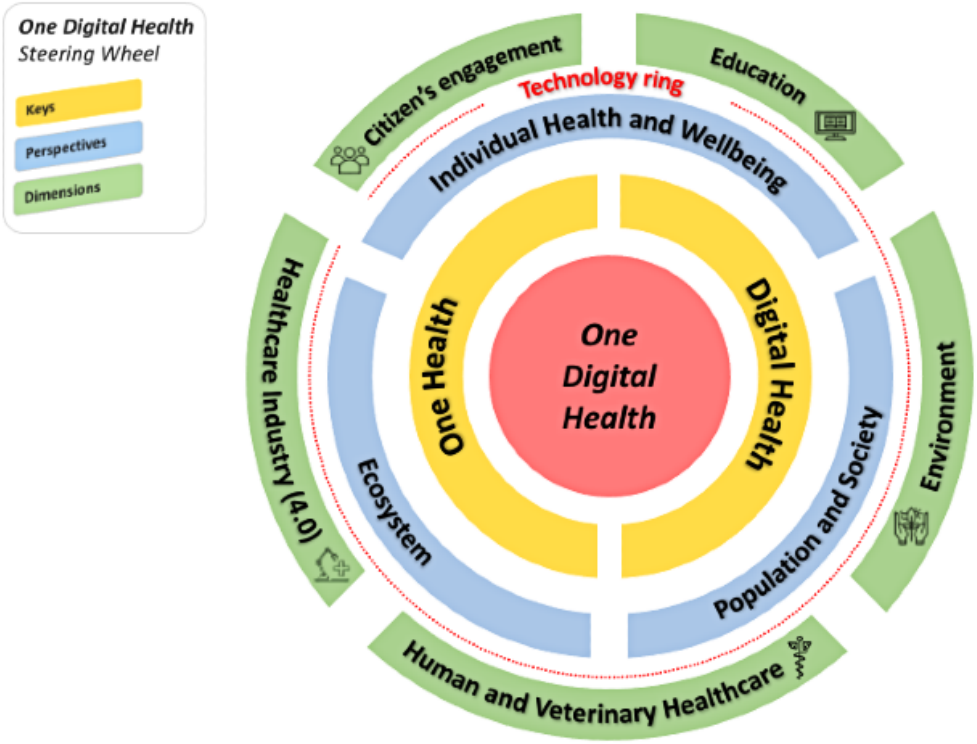
What is ODH?

ODH aims to facilitate and improve collaboration among practitioners in One Health and digital health communities. This collaboration will allow both communities to benefit from efficient interactions over time and the delivery of near–real-time, data-driven contributions to systems medicine [16] and systems ecology [17]. This will also allow citizens to engage with their individual health and well-being.

ODH is comprised of three intertwined levels (ie, two enabling keys, three perspectives, and five dimensions), wherein digital technology acts a catalyst. These are described in the following sections.

Figure 1 shows our conceptualization of the ODH framework, which we named the ODH Steering Wheel.

Figure 1. The One Digital Health Steering Wheel conceptual framework.



The Two Enabling Keys: One Health and Digital Health

It is widely acknowledged that the concepts of health data, information, and knowledge refer to a person's health and medical history [18]. Health informatics refers to the management (ie, collection, aggregation, analysis, and interpretation) of health data, information, and knowledge (ie, those related to patient care) that are made available through the deployment of digital tools [19]. Many scholars have contributed to the expansion of these original core concepts. Such contributions include (1) increasing technology adoption and performance; (2) enhancing technology safety, quality, effectiveness; and (3) improving the efficiency of care. One Health and digital health researchers seek to make their fields more comprehensive and inclusive, by studying research fields that are constantly being improved due to the development of health informatics [20]. As a result, the field of biomedical and health informatics (BMHI) "pursues the effective uses of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, motivated by efforts to improve human health"; and "investigates and supports reasoning, modeling, simulation, experimentation, and translation across the spectrum from molecules to populations, dealing with a variety of biological systems" [21]. In this sense, BMHI reflects manifold purposes and benefits all professionals in the health care field [22].

During the early 2000s, the internet was integrated into the day-to-day life of the general public. At the same time, eHealth emerged as a novel concept. eHealth has expanded the focus of medical informatics to include clinical information systems [23]. This has allowed the field to actively acknowledge health service customers, which are the central focus of the system. Thus, eHealth has majorly emphasized the importance of consumer health informatics in the current century [24].

eHealth-related concepts have been defined [24] and disseminated over more than 20 years [25]. Nonetheless, the massive process of digitalizing health services (ie, the process we are currently witnessing) has brought renewed focus to economics. This focus has manifested as the use of the term "digital health" [26]. This means that even though the terms "electronic something" and "digital something" are considered synonyms, the former seemingly refers to "electromagnetic stuff" from the "old times" (ie,1990s to the beginning of 2000s), whereas the latter sounds more like "advanced technologies" that fully support the daily practices of health and care. Thus, the term "digital health" is catchier than "electronic health," which can refer to communities that have not fully integrated technology [26]. In other words, no considerable differences seem to exist between eHealth and digital health, as no paradigm shifts exist between them. Yet, digital health can be seen as a refined version of eHealth, since digital health explicitly accounts for the deployment of health data, information, knowledge, and decision support systems via digital technologies, in a context that features life and environmental aspects [23]. Digital health therefore entails (1) ubiquitously collecting and storing data, information, and knowledge to efficiently deliver health care; (2) making health and medicine more personalized and precise; (3) pursuing goals that are related to health promotion, well-being, and efficient self-management; (4) considering the economic dimensions of the services tendered.

The deployment of digital health interventions requires expertise from health practitioners; health researchers; and scientists in the fields of engineering, social sciences, public health, health economics, and data management [21]. It is for these reasons

that we use the term "digital health" instead of "eHealth" in this viewpoint paper.

Concepts such as health, well-being, individual, population, living quality, and biological systems stand at the core of One Health practitioners' collaborative, multisectoral, and transdisciplinary approach. One Health practitioners also aim to achieve optimal health and well-being outcomes at the local, regional, national, and global levels, by recognizing the interconnections between people, animals, plants, and their ecosystems [27-29]. People's rising interest in the manifold competence areas of One Health have allowed them to witness the development of innovative strands of research, such as digital epidemiology and public health infodemiology research [30,31]. These new disciplines aim to improve people's understanding of health risks, effective management, and policy decisions [32,33]. In this spirit, One Health informatics has been proposed as an approach for bringing these research topics together during public health and life sciences research. The key themes of One Health informatics are deploying big data analytics for supporting and improving public health and medical research, and addressing issues that are related to biodiversity control, disaster management, and disease monitoring (eg, the surveillance of zoonoses for determining early warning signs, prevention methods, and control methods).

One Health practitioners worldwide have been calling for engagement at a much broader, ecosystem-level scale. Systemic and cultural transformations are necessary, as they allow us to appreciate how disruptive technologies play a big role in the implementation of new practice models, which encourages new ways of thinking about health, care, and wellness. Thus, the profound impact of disruptive technologies is expected to shape the future of health systems. This is why the very idea of "health systems" should be reconsidered [34,35]. We propose that the One Health informatics frame, which includes the concepts of living, society, and digital health intervention–related expertise, should be included in One Health's widened framework, which looks at the inextricable interconnectedness between humans, animals, and the environment [9,10]. The following aspects should be included: (1) the delivery of health care in a syndemic scenario [36]; (2) the digital transformation of human and animal health data (ie, including the concept of animal welfare) [37]; and (3) digital nature conservation (ie, biodiversity and wildlife conservation, food security, antimicrobial resistance, climate change, etc) by means of digital technology–based interventions [38].

A wide range of scientific disciplines has been called upon to work together to address emerging issues that are related to the above aspects. This holistic set of aspects and expertise requirements can be brought together under the proposed ODH framework.

The Three Perspectives: The "Individual-Population-Ecosystem" Triad

The perspectives layer of the ODH looks at how the concept of "individual" is identified within that of "population" (ie, without distinguishing between human and animal), how a population impacts and interacts with an ecosystem, and how the ecosystem responds accordingly.

Individual Health Care and Well-being

The emerging awareness of shared risks that face animal and human populations (ie, a concept that was originally part of the One Health approach) fosters opportunities for leveraging large amounts of unexplored data sources on zoonoses to generate new information and knowledge. The scope of the ODH includes supporting rapid and more accurate methods for detecting disease trends, outbreaks, pathogens, and causes of emergence [8,39-41]. The recognized, paramount importance of collecting digital health data to strengthen our understanding highlights the need for frameworks and tools that support and acknowledge the shift from big data to smart data. Smart data involves innovative processes that are necessary for answering pivotal questions from the health sector [33]. Addressing these issues requires new, ecosystem-driven paradigms and transformative technologies that reinforce traditional surveillance systems' ability to prevent and control diseases. These innovations will further impact intersectoral coordination; link human and animal health data; allow for the flow of reliable information and knowledge; and promote the proper use of infrastructures, systems, and human resources for detecting outbreaks, thereby ensuring good health and well-being.

Population and Society

Discussing populations and their societal organization does not simply mean providing a generalization of concepts that are related to personalized health care and well-being. Although we are discussing individual-level, predictive, personalized, preventive, and participatory health care [42-44], shifting to the topic of population-level, unified perspectives requires us to account for the differences between individuals. Contrary to a one-size-fits-all approach, personalized health has to take into account an individual's variability in terms of genes, environments, digital health literacy, preferences, and lifestyles. This requires a systematic integration of knowledge (ie, knowledge about health care, veterinary care, agriculture, meteorology, climate change, environmental protection, and intelligence) [45] into novel decision-making processes. These processes include decisions about the prevention and management of all hazards and threats to public health, such as exposure to animal diseases with zoonotic potential, environmental exposures, drug adverse events, the rise of misinformation [46], instances of diseases, and disease outbreaks [47].

There is a lack of solid and recognized leadership. Therefore, establishing an ethical framework that is socially and culturally accepted will lead to thriving digital platforms that can develop and grow independently. Such platforms will evolve to deliver the needs of a business.

Innovative Ecosystem

Due to peoples' current understanding of the term "environment," many assume that "environment" refers to a systemic perspective that focuses on the deep, existing connections among the sky, water reserves, and soil (ie, the global environment or the environment of specific geographic

areas). In this viewpoint paper, the concept of environment refers to the consideration of all living and nonliving things in an ecosystem. This accounts for issues that are related to biodiversity conservation and the intimate links among the health, care, and well-being of all components in any given ecosystem [48].

The control and governance of environmental sustainability can be best approached by assessing ecosystem services that are capable of quantifying and valuing all the goods and services that are generated within the ecosystems themselves. Recently, such networks have been increasingly endowed with digital technologies such as (1) environmental management and monitoring information systems; (2) automated and scalable approaches for collecting, digitalizing, and assembling geocoded big data; and (3) information-fusion algorithms that use multiple data streams and clinical decision support algorithms that integrate population-based, public health–focused perspectives into outbreak detection–focused management systems [49-51].

Environment turns therefore out as "digital" in ecosystems that, under a digital vest, feature animals and humans, as well as software and robots, as autonomous and interacting agents to enhance decision support. The goals of digital ecosystem implementation include improving the efficiency of communication and avoiding peoples' dependence on centralized or distributed control dynamics [52]. Such digital biodiversity originates from common substrates of data, information, and knowledge that are now accessible, available, and able to be analyzed in novel and innovative ways. However, this puts us into a Copernican situation [53], which is a prerequisite to figuring out appropriate digital solutions (ie, methods for safeguarding digital biodiversity) that are based on people's intelligence and rationality. Such solutions are preferable over those that are gained through traditional artificial intelligence methods [52,54].

The Three Perspectives as a Whole

In the broad digital ecosystem, ODH interventions must support, improve, and lead to efficient end-to-end processes for predictive, personalized, preventive, and participatory health care [42]. The management of disruptive innovation is the catalyst required for making this leap forward [55]. This requires the systematic, continuous, and intelligent integration of big, smart, and multidimensional data into digital health technologies. Such technologies are used for detecting, monitoring, and tracking the origins and causes of emerging pathogens, disease outbreaks, and disease-related trends that affect both humans and animals in the ecosystem. Thus, a challenging key issue that ODH must face is collecting and managing critical indicators that are related to human and veterinary health care, education, citizen engagement, environmental observation, and the health care industry. However, these indicators are required for active innovation. The COVID-19 outbreak is an example of an event that has triggered digital innovation and revolution in health care (ie, including ODH) on a global scale; the pandemic has quickly strengthened the overall care domain [56].

The three perspectives in the "individual-population-ecosystem" triad need to be further assessed as individual perspectives and as a specific set of strongly interconnected dimensions. These aspects will be discussed in-depth in the next sections.

The Five Dimensions: The Essential Fabric of Interconnections

The dimensions layer provides an integrative view of ODH. This layer shows society's outlook on the three perspectives and, in turn, the two keys. ODH depends on citizens' engagement and their ability to support and contribute to health care organizations in a continuously forward-moving environment.

Education and Citizen Science

ODH should not be limited to higher education in health-related fields. Basic and transversal knowledge and an understanding of health and its digitalization must be provided by education programs for all levels and disciplines (ie, similar to how the COVID-19 pandemic needs to be understood at all levels). ODH is about more than supporting data management. ODH provides learning opportunities to citizens, so that they can protect and take care of themselves when they share personal data on social networks [57-59]. At its core, ODH teaches people how to differentiate between real and fake news, which are disseminated over the mass media and the internet.

The next generation of researchers and practitioners will be actively involved in redefining critical transition points for incorporating new knowledge into ODH, and constantly updating the underlying philosophy and the scientific, technological, and regulatory implications of ODH [60]. Fully appraising an ODH perspective is challenging due to its intrinsic complexity. Therefore, researchers, practitioners, and citizens need to engage in science and cocreate sound implications for health and well-being. The core challenges of ODH include (1) establishing a systemic and integrated understanding of the health and wellness of humans and animals in their common ecosystem; and (2) establishing how digital systems may support and improve the health and wellness of humans and animals. Therefore, it is necessary to provide learners with project-based learning and case-based learning, which are well-known methods that allow learners to autonomously develop their skills and form learning communities [61-64].

Citizen Engagement

Supporting digital health literacy by actively encouraging everyone (ie, from children to elderly people) to engage with personal health, public health, and environmental monitoring systems can increase citizens' awareness. Advancements in mobile and ubiquitous tracking, reporting, and follow-up apps/web-based systems have provided opportunities for citizens to engage and collaborate with governmental, health, environmental, and ecological organizations [65]. Therefore, opportunities in which citizens use their electronic health data to manage personal health and share personal information require citizens who trust in health care delivery. The use of implemented opportunities depends on patient engagement, the direct benefits for users, and health care management organizations [66]. Smart cities and neighborhoods (ie, cities and neighborhoods that extensively use smartphones that can geolocate individuals and trace/track behaviors) [67] can be

intrusive sources of information. The act of intrusively obtaining information has introduced ethical dilemmas, challenged citizens' trust, and disrupted societal norms. The European General Data Protection Regulation has been implemented to reduce the amount of excess data in the overall data sphere. Nevertheless, in cases of force majeure, intrusive approaches that are implemented during pandemics outbreaks (eg, the COVID-19 outbreak) may help public health decision makers understand how to develop efficient policies that are based on real-world and near–real-time data [68]. Therefore, policy decision makers may decide to passively involve the population by allowing the use and analysis of data on individuals, mass movements, and individuals' (ie, humans' and animals') engagement with and access to certain facilities. This would mean having citizens engage with complex policy making over time (ie, for policies that are related to ethics, regulation, decisions on big smart data, and the shaping of social norms) [69]. From an ODH point of view, collaboration and transparency are essential for enhancing health systems' and government authorities' times to response during disruptive and potentially major events [45]. For example, in Taiwan, efficient smart contact tracing analyses were conducted to support an automated alert messaging system, which informed citizens about who required quarantine and isolation after coming into contact with the potentially infected passengers of the Diamond Princess cruise ship [70]. Furthermore, the European eHealth Network established the cross-border directive on patient's rights to cross-border health services (ie, article 14), and recently developed guidelines for the interoperability of contact tracing apps [71]. These examples show how ODH views the health-related and environment-related movements of individuals. However, several countries have discontinued COVID-19 contact tracing apps after data protectorate organizations considered these apps to be too intrusive [72].

Human and Veterinary Health Care

ODH aims to answer complex questions about zoonotic disease follow-ups by assessing public health impacts, surveilling human and animal health, and conducting related risk assessments and management processes [73,74]. Recently, health care delivery services have focused on providing services to clusters of patients [75]. The timely identification of health care customers, efficient service provision, and the continuous improvement of care quality accelerate the digital transformations of human and veterinary health care systems. In the context of the COVID-19 pandemic, environmental factors (eg, weather, pollution, and food) and animal health (eg, pets, livestock, and food) [76] influence human health, care, and well-being.

Digitalization and the increasing use of mobile and ubiquitous health/wellness apps and platforms have allowed for the assessment of patients (ie, human and animal patients) via dynamic, near–real-time methods that use patient-generated and participatory data. These methods enhance personalized health delivery services, such as preventive recommendations and predictive alert systems [77]. This new time- and space-oriented paradigm recalls connected health model features. It also allows us to view teleservices (ie, digital health services) as a full component of integrated care for health care customers and citizens who share their data to support the development of new health care practices and guidelines.

An example of connected health tools is SMS text messaging services, which are used to maintain patients' involvement in their treatments (eg, vaccinations, rehabilitation, and medication renewals) by improving their adherence to therapy [75,78,79]. In terms of pets or livestock, owners can use connected health tools to stay involved in the health surveillance and follow-up of their animals [80]. In terms of the ODH framework, connected health tools support efforts for increasing global health security, as these tools help with collecting near–real-time feedback from populations and health care organizations [81].

Digital Health Care Transformation 4.0

The health care system that deals with humans and animals is comprised of (1) health care service providers and insurers; (2) medical equipment and pharmaceutical product producers, sellers, deliverers, and maintenance service providers; (3) regulatory agencies and standardization organizations; and (4) health care customers, patients, and caregivers.

The entire health care sector is driven by digitalization and interconnectivity, which make up the foundation of new automation paradigms. The main challenges of the new health care framework include managing, collecting, storing, archiving, and analyzing a wide range of real-time data that are made available by any kind of organization and any kind of connected system. The leading goals of the current industrial revolution (ie, Industry 4.0) [82] are to enhance the automation and connectivity of systems, by increasing the interoperability and flexibility of systems and allowing for decentralized, real-time data collection and storage. The objective of Industry 4.0 is to use findable, accessible, interoperable, reusable, ethical, and revisable (FAIRER) data [83]. By adopting a FAIRER model, health care will benefit from greater traceability, flexibility, adaptability, and efficiency in health delivery processes [84]. As the health industry continues to adopt the ODH paradigm, a strong dependence on smart data means that standardization and interoperability framework issues are more critical than ever. A large number of systems need to be interconnected to collect real-time data that are generated by all parts of the health care system. The Health Level 7 Fast Healthcare Interoperability Resources standard [85] has been quite successful in this regard, as it focuses on the use of application programming interfaces within and across health/care sectors to improve a wide range of areas (ie, interconnectedness to contractual agreements).

The global health care industry has used the same technologies as those used by Industry 4.0 in the digital era [86], to achieve the use of the well-known "4.0" suffix. Telemonitoring systems use technologies that were initially designed for Industry 4.0. These technologies have been extensively used in human medicine [87,88]. Recently, their use has expanded as a result of the COVID-19 outbreak [89], as teleconsultation technology use has become the new norm in medical care. In terms of ODH, the same approach has been used by veterinarians. They can use Industry 4.0–related technologies to receive health and wellness information from wearable devices or implanted sensors that provide data on physiological signs. Veterinarians can also use these technologies to receive third-party, pet-related

information [90]. The integration of these systems into pet health care may provide a better understanding of factors that affect pets' health and pet owners' benefits. Moreover, data on changes in a pet's health can be used as pet owner health indicators [91] and environmental quality indicators [92].

3D printing is an emerging dynamic field in health care industries [93], including therapeutic industries (eg, printing prosthetics implants), human and veterinary care industries (eg, 3D printing in dental surgery), and educational industries [64]. The use of this on-demand technology supports the provision of personalized therapies and medical products to patients who are directly involved with health care practitioners. 3D printing has been shown to reduce the time to treatment by removing the need to travel back and forth between production laboratories and clinics.

Environment

Environmental monitoring is a critical dimension of community-oriented management in day-to-day situations or disruptive situations, such as hazardous material incidents or disasters (eg, the COVID-19 pandemic) [94,95]. The digital landscape has allowed theoretical and laboratory-limited developments in automation and artificial intelligence (ie, those in the last several decades) to be put into practice [96]. Discussing the environment dimension of ODH means having to discuss the "human-nature-pollution" relationship and its role in the smart city paradigm. People who live in high-density urban areas have been increasingly using technological systems to report unusual events, thereby improving citizen engagement [97,98]. Furthermore, these people have been increasingly using green technologies, which reduce the negative impact of certain actions on the environment (eg, carbon dioxide emissions, car traffic, and waste production) and health [99].

With regard to the ODH, this environmental revolution means more than using smartphone and Internet-of-Things technologies for behavior, wellness, and health monitoring purposes. This revolution is a consequence of the Industry 4.0 revolution [100], which aims to enable citizens and their communities to interact with complex, but easy-to-use, digital systems. However, the Industry 4.0 revolution has different issues, which are mainly related to development costs, monitoring costs, implementation costs, maintenance costs, and the understanding and expectations of users (eg, citizens and decision-makers) [101].

The environment dimension is strongly related to the human and veterinary health care, One Health, and ODH. Wild and domestic animals are used as a part of surveillance platforms that provide early warnings about health hazards, which may impact the whole ecosystem. This is an example of why the environment, animals, and humans are part of the One Health "individual-population-ecosystem" triad [102]. Environment systems must be able to integrate, and be integrated into, health care management systems, to enrich data, information, and knowledge; actively support decision-making processes about environmental exposures and health risks; and disseminate efficient recommendations with the right communication tools [103]. This ODH dimension also relates to smart environment platforms that provide information to smart homes to monitor and deliver personalized services to older adults [104].

Digital Technologies as a Catalyst for the Integration of Keys, Perspectives, and Dimensions

Technology as a Catalyst for "ODHness"

The main aims of digital health are improving health, care, wellness, and public health through BMHI research [22]; digital health expands these concepts by allowing for the consideration of digital consumers who use a wide range of smart-devices, connected medical equipment, or connected wellness equipment. Digital health also encompasses other digital technologies for health, such as the Internet of Things, artificial intelligence, big data, and robotics. These technologies have become a full component of day-to-day life, [84] and in some cases, health-related crisis management [46,70]. Healthy individuals, chronic patients, and health management organizations have been increasingly using automated reminder systems, smartphone apps, and health monitoring wearables [75,105,106]. In terms of ODH, technology serves as the catalyst for digitizing One Health information for ODH; advanced technological innovations have emerged to improve citizen engagement and empower future health ecosystems (Figure 1).

In health crises like the COVID-19 pandemic, the applications of digital technologies (eg, managing, preparing for, and mitigating crises) include the planning and scheduling of response and recovery processes. Digital technologies use surveillance, contact tracing, contact tracking, testing, confinement, and other health and care methods when necessary (eg, after ethical and legal implications have been considered) [107]. A similar level of attention is necessary for effectively tracking veterinary information that needs to be extracted from vast amounts of data. This process involves the integration of animal medical data into real-time information systems that are dedicated to supporting public health [108]. Thus, the role of veterinarians and veterinary informatics enhances that of BMHI (ie, protecting health care), as veterinarians have been called on to help deal with the increase amounts of dynamic data on emerging and reemerging infectious diseases. Local and supranational indicators of health have been used to identify global-level risk factors and causes of health problems that arise at the human, animal, and environmental levels, which need to be taken into account [109,110].

According to the WHO vision, establishing an interoperable digital health ecosystem that is capable of seamless, secure health data exchange and processing is crucial for combating pandemic outbreaks. Developing infrastructures and apps that allow the use of health data to manage adverse events is within the scope of achieving the 17 health-related sustainable development goals that were issued in the WHO global action plan [111]. We therefore introduced the concept of digital One Health interventions [112] in this viewpoint paper as a set of digital functionalities that should be designed and deployed to (1) support specific initiatives that address human, animal, and environmental systems' needs and challenges; and (2) assess, study, and collect data on these systems' expected outcomes, unexpected outcomes, and effects [36]. This related to the

selection of timely metrics for the outcomes of multicriteria decision analyses.

These digital functionalities can be also called "digitalities," as they account for how technology has been embedded into human experiences and humans' daily lives. This has resulted in core issues in social and cultural anthropology (ie, digital humanities) [112]. Digitalities also account for how humans affect animals' daily lives, animal health issues, and human-animal relations. These digitalities aim to increase animal welfare (ie, digital animalities) [113]. Furthermore, digitalities aim to manage the complex web of human, animal, and environmental interconnections, by improving environmental governance (ie, digital environmentalities) [114,115].

The delivery of digital One Health interventions and the mechanisms of the impacts and contextual factors of these interventions can be referred to as "ODHness." This term is based on the combined digitalities of the three overarching, complementary perspectives. These digitalities apply to the five ODH dimensions and their constituent subcategories. Therefore, the technology ring in Figure 1 serves as the connection between the ODHness concept and relevant digitalities.

Figure 2 shows several topics that are relevant to each of the three digitalities (ie, the three colored areas inside the technology ring). Each topic is characterized by each portion within the corresponding colored area; the width of these areas depends on the number and the nature of the area's related subtopics. It should be noted that these subtopics have not been clearly represented in the figure to ensure that people comprehend the ODH scheme. In terms of digitality, a specific metric or set of metrics must be used to evaluate the levels of development, use, contribution for each digital One Health intervention topic. These levels are represented with black dots for each subdomain, connected with each other via dashed lines. Since interactions between topics from different digitalities are possible, the edges that connect the different areas are represented with different colors. In addition, "trialities" (ie, triangles that connect all three digitalities) are represented via thick lines. Furthermore, each edge is assigned an appropriate weight. This of course means that harmonization must exist between the implemented metrics. Harmonization is required to fulfill the needs of a multicriteria decision analysis.

The technology ring in Figure 2 encompasses the set of digital functionalities that the ODH framework relies on. Therefore, technology acts as the catalyst that connects and unifies the ODH disciplines. ODHness is represented by the technology ring's "center of mass" (ie, the red dot in Figures 2 and 3). More specifically, the "center of mass" is the unique point where the weighted relative position of the distributed mass (ie, the edges) equals 0. In other words, this means that a simple set of digital technologies (ie, those related to the human, animal, and environmental fields) is not capable of providing timely solutions in an interdisciplinary context, unless these technologies act as the catalyst for balancing the different perspectives, needs, and interests of ODH. In Figure 2, we attempted to highlight the "human-animal-environment" computer interaction research area, which includes technology designs that support animals in different contexts, and the development of user-centered approaches for designing technologies that are intended for more-than-human animals [116].

Examples of possible pandemic-related interactions within the technology ring are shown in Figure 3. Several of the main topics of each digitality are shown. The levels of development, use, and contribution for each digital One Health intervention are also shown. For convenience, the ODHness point is represented by the center of the technology ring.

Figure 2. An example of the interactions within One Digital Health, which are based on the use of technology as a catalyst.

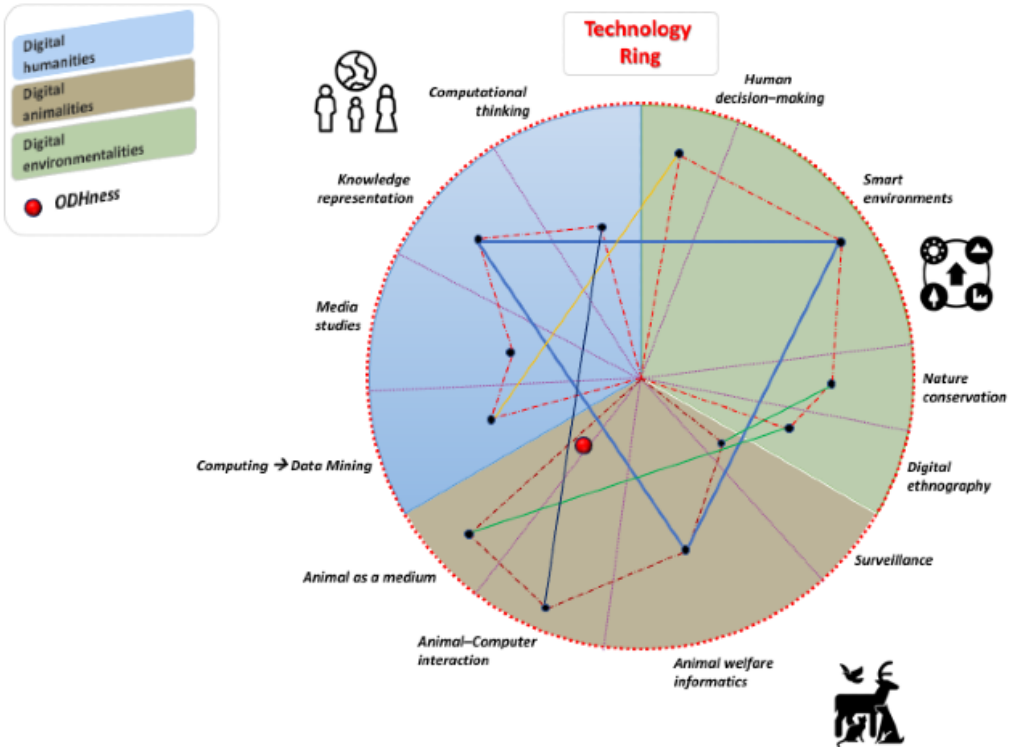
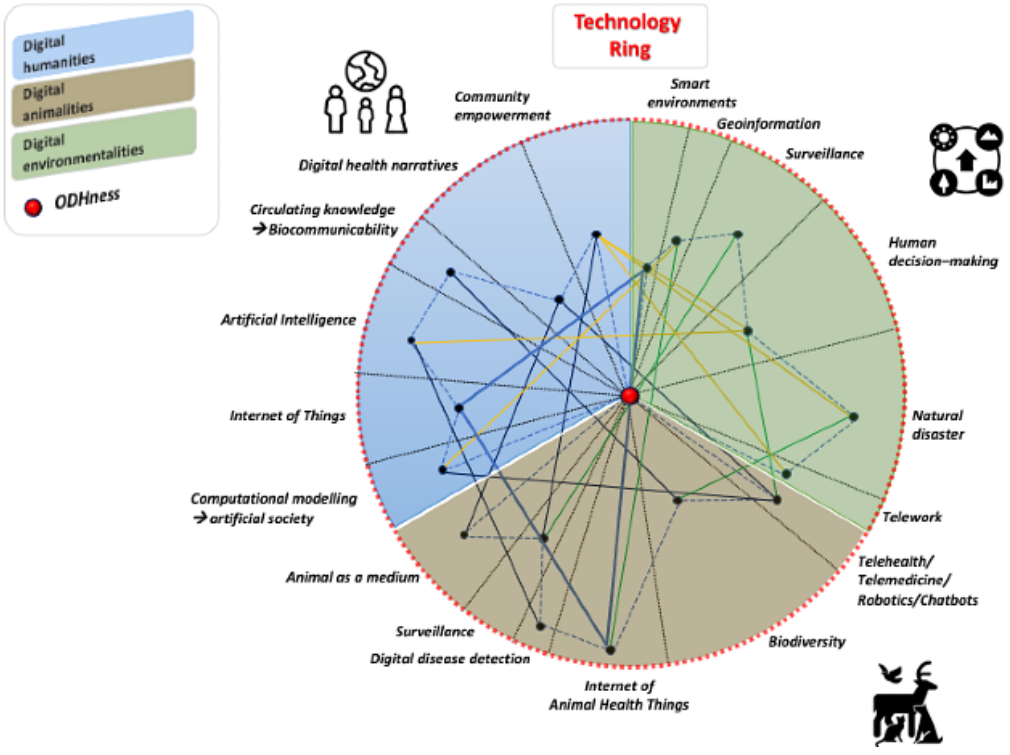


Figure 3. Pandemic-related "ODHness" interactions within the technology ring.



Discussion and Conclusions

ODH is a framework that integrates key dimensions from two established fields—digital health and One Health. The latter focuses on the interconnectedness of human, animal, and environmental health research. The former focuses on providing tools and domain expertise to support health practitioners in their administrative (eg, consultation scheduling), clinical (eg, necessary tests, drug-drug interaction alerts, and specific follow-up reminders) and scientific (eg, disease prevalence analysis, data collection, and reporting) activities. The digital health field also encompasses citizens, who are the consumers of health services (eg, healthy individuals, sick individuals, relatives who assist sick individuals in day-to-day life, and insurers).

ODH analyzes the health data and information of digital health ecosystem components, because the intents, processes, and products of these components (ie, health, care, and wellness) are integrated into health ecosystems. Moreover, ODH analyzes three perspectives. These perspectives include three different areas that deal with the growing scientific evidence and perspectives that come with technologies that are meant to support health, care, and well-being activities in separate ecosystems. Furthermore, this unified framework for future health ecosystems analyzes the impact of five dimensions (eg, education provides short-, mid-, and long-term learning opportunities for improving citizens' engagement with human and veterinary health care services; environment monitoring; and health care industry–developed, interoperable early warning systems).

The COVID-19 pandemic has increased the world's understanding of the close relationships among the environment, animals, and humans [117]. Therefore, it is more critical than ever to effectively manage the pandemic by assessing the innovative two keys, three perspectives, and five dimensions in the unifying ODH framework. It is also important to assess the pandemic's impact on health ecosystems by facilitating systematic collaboration among animal and environmental scientists, health care practitioners, citizens, governments, academics, and industrial manufacturers.

The next step in developing and testing the ODH framework will be performed by the proposed EFMI ODH working group. The proposed EFMI ODH working group will elaborate upon the ODH framework, validate the methodology, and evaluate the ODH. They will also examine specific ODH scenarios in-depth and analyze the relationships between the perspectives in the ODH framework and BMHI dimensions (ie, data, information, and knowledge). Analyzing these relationships is expected to yield new insights and support an integrative, systemic, and syndemic decision-making process that can be assessed from a multidisciplinary perspective [117,118]. We expect that this process will be instrumental in shaping future health ecosystems, providing novel learning opportunities for citizens, and educating the next generations of BMHI practitioners.

Conflicts of Interest

None declared.

References

1. Ellis JL, Jacobs M, Dijkstra J, van Laar H, Cant JP, Tulpan D, et al. Review: Synergy between mechanistic modelling and data-driven models for modern animal production systems in the era of big data. Animal 2020 Aug;14(S2):s223-s237 [FREE Full text] [doi: 10.1017/S1751731120000312] [Medline: 32141423]
2. Todde V, Giuliani A. Big data. A briefing. Ann Ist Super Sanita 2018;54(3):174-175. [doi: 10.4415/ANN_18_03_02] [Medline: 30284542]
3. Hu H, Galea S, Rosella L, Henry D. Big data and population health: focusing on the health impacts of the social, physical, and economic environment. Epidemiology 2017 Nov;28(6):759-762. [doi: 10.1097/EDE.0000000000000711] [Medline: 28682850]
4. Mackenzie JS, Jeggo M. The one health approach-why is it so important? Trop Med Infect Dis 2019 May 31;4(2):88 [FREE Full text] [doi: 10.3390/tropicalmed4020088] [Medline: 31159338]
5. Sustainable development goals. Food and Agricultural Organization of the United Nations. URL: http://www.fao.org/sustainable-development-goals/goals/goal-3/en/ [accessed 2020-07-05]
6. Gostin LO, Friedman EA. The sustainable development goals: One-Health in the world's development agenda. JAMA 2015 Dec;314(24):2621-2622. [doi: 10.1001/jama.2015.16281] [Medline: 26660329]
7. Zinsstag J, Schelling E, Waltner-Toews D, Tanner M. From "one medicine" to "one health" and systemic approaches to health and well-being. Prev Vet Med 2011 Sep 01;101(3-4):148-156 [FREE Full text] [doi: 10.1016/j.prevetmed.2010.07.003] [Medline: 20832879]
8. Rosa W. Chapter 34: One mind, one health, one planet - A pledge to planetary citizenship. In: A New Era in Global Health. New York City: Springer Publishing Company; 2017.
9. Osterhaus A. Welcome to One Health Outlook. One Health Outlook 2019 Nov 27;1(1):1-2. [doi: 10.1186/s42522-019-0005-y]
10. Ashleigh C. Chapter 11: Visualising One Health. In: One Planet, One Health. Sydney: Sydney University Press; 2019.
11. Swei A, Couper LI, Coffey LL, Kapan D, Bennett S. Patterns, drivers, and challenges of vector-borne disease emergence. Vector Borne Zoonotic Dis 2020 Mar;20(3):159-170. [doi: 10.1089/vbz.2018.2432] [Medline: 31800374]
12. Yamane I. Epidemics of emerging animal diseases and food-borne infection problems over the last 5 years in Japan. Ann N Y Acad Sci 2006 Oct;1081:30-38. [doi: 10.1196/annals.1373.003] [Medline: 17135492]
13. de Lusignan S, Bernal JL, Zambon M, Akinyemi O, Amirthalingam G, Andrews N, et al. Emergence of a novel coronavirus (COVID-19): Protocol for extending surveillance used by the Royal College of General Practitioners Research and Surveillance Centre and Public Health England. JMIR Public Health Surveill 2020 Apr 02;6(2):e18606 [FREE Full text] [doi: 10.2196/18606] [Medline: 32240095]
14. Jin Y, Yang H, Ji W, Wu W, Chen S, Zhang W, et al. Virology, epidemiology, pathogenesis, and control of COVID-19. Viruses 2020 Mar 27;12(4):372 [FREE Full text] [doi: 10.3390/v12040372] [Medline: 32230900]

15. Lovis C. Digital health: A science at crossroads. Int J Med Inform 2018 Feb;110:108-110. [doi: 10.1016/j.ijmedinf.2017.12.006] [Medline: 29331249]
16. Comte B, Baumbach J, Benis A, Basílio J, Debeljak N, Flobak Å, et al. Network and systems medicine: Position paper of the European Collaboration on Science and Technology Action on Open Multiscale Systems Medicine. Netw Syst Med 2020 Jul 06;3(1):67-90 [FREE Full text] [doi: 10.1089/nsm.2020.0004] [Medline: 32954378]
17. Borrett SR, Fath BD, Whipple SJ. Introduction to the special issue “Systems Ecology: A Network Perspective and Retrospective”. Ecological Modelling 2014 Dec 10;293:1-3. [doi: 10.1016/j.ecolmodel.2014.10.005]
18. Vest JR, Gamm LD. Health information exchange: persistent challenges and new strategies. J Am Med Inform Assoc 2010;17(3):288-294 [FREE Full text] [doi: 10.1136/jamia.2010.003673] [Medline: 20442146]
19. Haux R. Medical informatics: past, present, future. Int J Med Inform 2010 Sep;79(9):599-610. [doi: 10.1016/j.ijmedinf.2010.06.003] [Medline: 20615752]
20. Agarwal R, Gao G, DesRoches C, Jha AK. Research commentary—The digital transformation of healthcare: current status and the road ahead. Inf Syst Res 2010;21(4):661-1010. [doi: 10.1287/isre.1100.0327]
21. What is informatics? American Medical Informatics Association. URL: https://www.amia.org/fact-sheets/what-informatics [accessed 2020-05-26]
22. Mantas J, Ammenwerth E, Demiris G, Hasman A, Haux R, Hersh W, IMIA Recommendations on Education Task Force. Recommendations of the International Medical Informatics Association (IMIA) on education in biomedical and health informatics. First revision. Methods Inf Med 2010 Jan 07;49(2):105-120. [doi: 10.3414/ME5119] [Medline: 20054502]
23. Khillar S. Difference between digital health and health informatics. Difference Between. URL: http://www.differencebetween.net/technology/difference-between-digital-health-and-health-informatics/ [accessed 2020-06-26]
24. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
25. Eysenbach G. Celebrating 20 years of open access and innovation at JMIR Publications. J Med Internet Res 2019 Dec 23;21(12):e17578 [FREE Full text] [doi: 10.2196/17578] [Medline: 31868653]
26. Ahmadvand A, Kavanagh D, Clark M, Drennan J, Nissen L. Trends and visibility of "Digital Health" as a keyword in articles by JMIR Publications in the new millennium: Bibliographic-bibliometric analysis. J Med Internet Res 2019 Dec 19;21(12):e10477 [FREE Full text] [doi: 10.2196/10477] [Medline: 31855190]
27. What is One Health? One Health Commission. URL: https://www.onehealthcommission.org/en/why_one_health/what_is_one_health/ [accessed 2020-06-26]
28. One Health. World Organisation for Animal Health. URL: https://www.oie.int/en/for-the-media/onehealth/ [accessed 2020-06-26]
29. One Health. World Health Organization. 2017 Sep 21. URL: https://www.who.int/news-room/q-a-detail/one-health [accessed 2020-06-29]
30. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
31. Mavragani A. Infodemiology and infoveillance: Scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
32. Ossebaard HC. One health informatics. In: WWW '14 Companion: Proceedings of the 23rd International Conference on World Wide Web. 2014 Apr Presented at: WWW '14: 23rd International World Wide Web Conference; April 2014; Seoul, Korea. [doi: 10.1145/2567948.2579274]
33. VanderWaal K, Morrison RB, Neuhauser C, Vilalta C, Perez AM. Translating big data into smart data for veterinary epidemiology. Front Vet Sci 2017 Jul 17;4:110 [FREE Full text] [doi: 10.3389/fvets.2017.00110] [Medline: 28770216]
34. Meskó B, Drobni Z, Bényei É, Gergely B, Győrffy Z. Digital health is a cultural transformation of traditional healthcare. Mhealth 2017 Sep 14;3:38 [FREE Full text] [doi: 10.21037/mhealth.2017.08.07] [Medline: 29184890]
35. Pita-Barros P, Bourek A, Brouwer W, Lehtonen L. Assessing the impact of digital transformation of health services. Report of the EXPH (Expert panel on effective ways of investing in health). Informační systém Masarykovy univerzity. URL: https://is.muni.cz/publication/1621656/022_digitaltransformation_en.pdf [accessed 2021-01-27]
36. Rock M, Buntain BJ, Hatfield JM, Hallgrímsson B. Animal-human connections, "one health," and the syndemic approach to prevention. Soc Sci Med 2009 Mar;68(6):991-995. [doi: 10.1016/j.socscimed.2008.12.047] [Medline: 19157669]
37. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: Key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
38. Arts K, van der Wal R, Adams WM. Digital technology and the conservation of nature. Ambio 2015 Nov;44 Suppl 4:661-673 [FREE Full text] [doi: 10.1007/s13280-015-0705-1] [Medline: 26508352]
39. Asokan GV, Asokan V. Leveraging "big data" to enhance the effectiveness of "one health" in an era of health informatics. J Epidemiol Glob Health 2015 Dec;5(4):311-314 [FREE Full text] [doi: 10.1016/j.jegh.2015.02.001] [Medline: 25747185]

40. Asokan GV, Kasimanickam RK. Emerging infectious diseases, antimicrobial resistance and millennium development goals: Resolving the challenges through One Health. Cent Asian J Glob Health 2013 Oct 01;2(2):76 [FREE Full text] [doi: 10.5195/cajgh.2013.76] [Medline: 29755885]
41. Ji X, Cappellari P, Chun SA, Geller J. Leveraging social data for health care behavior analytics. In: Engineering the Web in the Big Data Era. 2015 Presented at: International Conference on Web Engineering; June 9-12, 2020; Helsinki, Finland p. 667-670. [doi: 10.1007/978-3-319-19890-3_53]
42. Alonso SG, de la Torre Díez I, Zapiraín BG. Predictive, personalized, preventive and participatory (4P) medicine applied to telemedicine and eHealth in the literature. J Med Syst 2019 Apr 12;43(5):140. [doi: 10.1007/s10916-019-1279-4] [Medline: 30976942]
43. Flores M, Glusman G, Brogaard K, Price ND, Hood L. P4 medicine: how systems medicine will transform the healthcare sector and society. Per Med 2013;10(6):565-576 [FREE Full text] [doi: 10.2217/pme.13.57] [Medline: 25342952]
44. Sagner M, McNeil A, Puska P, Auffray C, Price ND, Hood L, et al. The P4 health spectrum - A predictive, preventive, personalized and participatory continuum for promoting healthspan. Prog Cardiovasc Dis 2017;59(5):506-521. [doi: 10.1016/j.pcad.2016.08.002] [Medline: 27546358]
45. Lai Y, Yeung W, Celi LA. Urban intelligence for pandemic response: Viewpoint. JMIR Public Health Surveill 2020 Apr 14;6(2):e18873 [FREE Full text] [doi: 10.2196/18873] [Medline: 32248145]
46. Eysenbach G. How to fight an infodemic: The four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
47. Tamburis O, Giannino F, D'Arco M, Tocchi A, Esposito C, Di Fiore G, et al. A night at the OPERA: A conceptual framework for an integrated distributed sensor network-based system to figure out safety protocols for animals under risk of fire. Sensors (Basel) 2020 Apr 29;20(9):2538 [FREE Full text] [doi: 10.3390/s20092538] [Medline: 32365698]
48. Rook GA. Regulation of the immune system by biodiversity from the natural environment: an ecosystem service essential to health. Proc Natl Acad Sci U S A 2013 Nov 12;110(46):18360-18367 [FREE Full text] [doi: 10.1073/pnas.1313731110] [Medline: 24154724]
49. Kim Y, Platt U, Gu MB, Iwahashi H. Atmospheric and Biological Environmental Monitoring. Berlin, Germany: Springer Science & Business Media; 2009.
50. Karatzas KD. Participatory environmental sensing for quality of life information services. In: Information Technologies in Environmental Engineering. Berlin, Heidelberg: Springer; 2011.
51. Khaiter PA, Erechtchoukova MG. Sustainability Perspectives: Science, Policy and Practice: A Global View of Theories, Policies and Practice in Sustainable Development. Switzerland: Springer International Publishing; 2019:A.
52. Boley H, Chang E. Digital ecosystems: Principles and semantics. In: Institute of Electrical and Electronics Engineers Explore. 2007 Presented at: 2007 Inaugural IEEE-IES Digital EcoSystems and Technologies Conference; February 21-23, 2007; Cairns, Australia p. 398-403. [doi: 10.1109/DEST.2007.372005]
53. Mori AR, Mazzeo M, Mercurio G, Verbicaro R. Holistic health: predicting our data future (from inter-operability among systems to co-operability among people). Int J Med Inform 2013 Apr;82(4):e14-e28. [doi: 10.1016/j.ijmedinf.2012.09.003] [Medline: 23122923]
54. La Salle J, Williams KJ, Moritz C. Biodiversity analysis in the digital era. Philos Trans R Soc Lond B Biol Sci 2016 Sep 05;371(1702):20150337 [FREE Full text] [doi: 10.1098/rstb.2015.0337] [Medline: 27481789]
55. Omachonu VK. Healthcare Value Proposition. New York: Productivity Press; 2018.
56. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82. [doi: 10.1056/NEJMp2005835] [Medline: 32240581]
57. Kallur A, Albalbissi A, Carillo-Martin I, Boonpheng B, Kallur L, Kherallah Y, et al. Doctor YouTube's opinion on seasonal influenza: A critical appraisal of the information available to patients. Digit Health 2020 Feb 16;6:2055207620906968 [FREE Full text] [doi: 10.1177/2055207620906968] [Medline: 32110430]
58. Reddy K, Kearns M, Alvarez-Arango S, Carrillo-Martin I, Cuervo-Pardo N, Cuervo-Pardo L, et al. YouTube and food allergy: An appraisal of the educational quality of information. Pediatr Allergy Immunol 2018 Jun;29(4):410-416. [doi: 10.1111/pai.12885] [Medline: 29512839]
59. Pizzuti AG, Patel KH, McCreary EK, Heil E, Bland CM, Chinaeke E, et al. Healthcare practitioners' views of social media as an educational resource. PLoS One 2020 Feb 06;15(2):e0228372. [doi: 10.1371/journal.pone.0228372] [Medline: 32027686]
60. Kolokathi A, Hasman A, Chronaki C, Madsen I, Moen A, Randell R, et al. Education in biomedical and health informatics: A European perspective. Stud Health Technol Inform 2019 Aug 21;264:1951-1952. [doi: 10.3233/SHTI190729] [Medline: 31438423]
61. Benis A. Healthcare Informatics Project-Based Learning: An Example of a Technology Management Graduation Project Focusing on Veterinary Medicine. Stud Health Technol Inform 2018;255:267-271. [Medline: 30306950]
62. Ricci FL, Consorti F, Pecoraro F, Luzi D, Mingarelli V, Tamburis O. HIN - Health issue network as means to improve case-based learning in health sciences education. Stud Health Technol Inform 2018;255:262-266. [Medline: 30306949]

63. Ricci FL, Consorti F, Pecoraro F, Luzi D, Mingarelli V, Miotti S, et al. Understanding Petri Nets in Health Sciences Education: The Health Issue Network Perspective. Stud Health Technol Inform 2020 Jun 16;270:484-488. [doi: 10.3233/SHTI200207] [Medline: 32570431]
64. Benis A, Nelke SA, Winokur M. Upgrading industrial engineering and management curriculum to Industry 4.0. 2020 Presented at: IEEE International Conference on Industrial Technology (ICIT); February 26-28, 2020; Buenos Aires, Argentina p. 355-359. [doi: 10.1109/ICIT45562.2020.9067243]
65. Moen A. Citizens and health data - Untapped resource for telehealth? Stud Health Technol Inform 2018;254:63-69. [Medline: 30306958]
66. Nohr C, Wong MC, Turner P, Almond H, Parv L, Gilstad H, et al. Citizens' access to their digital health data in eleven countries - A comparative Study. Stud Health Technol Inform 2016;228:685-689. [Medline: 27577472]
67. Servick K. Cellphone tracking could help stem the spread of coronavirus. Is privacy the price? Science Magazine. 2020 Mar 22. URL: https://www.sciencemag.org/news/2020/03/cellphone-tracking-could-help-stem-spread-coronavirus-privacy-price [accessed 2021-01-27]
68. Malgieri G. Data protection and research: A vital challenge in the era of COVID-19 pandemic. Computer Law & Security Review 2020 Jul;37:105431 [FREE Full text] [doi: 10.1016/j.clsr.2020.105431]
69. Brenner M, O'Shea MP, Larkin P, Luzi D, Pecoraro F, Tamburis O, et al. Management and integration of care for children living with complex care needs at the acute-community interface in Europe. Lancet Child Adolesc Health 2018 Nov;2(11):822-831. [doi: 10.1016/S2352-4642(18)30272-4] [Medline: 30336896]
70. Chen CM, Jyan HW, Chien SC, Jen HH, Hsu CY, Lee PC, et al. Containing COVID-19 among 627,386 persons in contact with the Diamond Princess cruise ship passengers who disembarked in Taiwan: Big data analytics. J Med Internet Res 2020 May 05;22(5):e19540 [FREE Full text] [doi: 10.2196/19540] [Medline: 32353827]
71. Interoperability guidelines for approved contact tracing mobile applications in the EU. eHealth Network. URL: https://ec.europa.eu/health/sites/health/files/ehealth/docs/contacttracing_mobileapps_guidelines_en.pdf [accessed 2021-01-27]
72. Morley J, Cowls J, Taddeo M, Floridi L. Ethical guidelines for COVID-19 tracing apps. Nature 2020 Jun;582(7810):29-31. [doi: 10.1038/d41586-020-01578-0] [Medline: 32467596]
73. Gongal G, Ofrin RH. Animal welfare, One Health and emergency preparedness and response in the Asia-Pacific region. WHO South East Asia J Public Health 2020 Apr;9(1):50-51 [FREE Full text] [doi: 10.4103/2224-3151.282996] [Medline: 32341222]
74. Short KR, Richard M, Verhagen JH, van Riel D, Schrauwen EJA, van den Brand JMA, et al. One health, multiple challenges: The inter-species transmission of influenza A virus. One Health 2015 Dec 01;1:1-13 [FREE Full text] [doi: 10.1016/j.onehlt.2015.03.001] [Medline: 26309905]
75. Benis A, Barkan RB, Sela T, Harel N. Communication behavior changes between patients with diabetes and healthcare providers over 9 years: Retrospective cohort study. J Med Internet Res 2020 Aug 11;22(8):e17186 [FREE Full text] [doi: 10.2196/17186] [Medline: 32648555]
76. Wilkinson A. Food for thought: food systems, livestock futures and animal health. Anim Health Res Rev 2013 Dec;14(2):159-162. [doi: 10.1017/S146625231300008X] [Medline: 23988197]
77. Chronaki C, Stegwee R, Moen A. In search of a digital health compass to navigate the health system. Stud Health Technol Inform 2017;245:30-34. [Medline: 29295046]
78. Menzies R, Heron L, Lampard J, McMillan M, Joseph T, Chan J, et al. A randomised controlled trial of SMS messaging and calendar reminders to improve vaccination timeliness in infants. Vaccine 2020 Mar 30;38(15):3137-3142. [doi: 10.1016/j.vaccine.2020.02.045] [Medline: 32147296]
79. Strandbygaard U, Thomsen SF, Backer V. A daily SMS reminder increases adherence to asthma treatment: a three-month follow-up study. Respir Med 2010 Feb;104(2):166-171 [FREE Full text] [doi: 10.1016/j.rmed.2009.10.003] [Medline: 19854632]
80. Madder M, Walker JG, Van Rooyen J, Knobel D, Vandamme E, Berkvens D, et al. e-Surveillance in animal health: use and evaluation of mobile tools. Parasitology 2012 Dec;139(14):1831-1842. [doi: 10.1017/S0031182012000571] [Medline: 22717001]
81. Osterhaus ADME, Vanlangendonck C, Barbeschi M, Bruschke CJM, Christensen R, Daszak P, et al. Make science evolve into a One Health approach to improve health and security: a white paper. One Health Outlook 2020;2(1):6. [doi: 10.1186/s42522-019-0009-7] [Medline: 32835168]
82. Bonilla SH, Silva HRO, Silva MT, Gonçalves RF, Sacomano JB. Industry 4.0 and sustainability implications: A scenario-based analysis of the impacts and challenges. Sustainability 2018 Oct 17;10(10):3740 [FREE Full text] [doi: 10.3390/su10103740]
83. Holub P, Kohlmayer F, Prasser F, Mayrhofer MT, Schlünder I, Martin GM, et al. Enhancing reuse of data and biological material in medical research: From FAIR to FAIR-Health. Biopreserv Biobank 2018 Apr;16(2):97-105 [FREE Full text] [doi: 10.1089/bio.2017.0110] [Medline: 29359962]
84. Chute C, French T. Introducing Care 4.0: An integrated care paradigm built on Industry 4.0 capabilities. Int J Environ Res Public Health 2019 Jun 25;16(12):2247 [FREE Full text] [doi: 10.3390/ijerph16122247] [Medline: 31242687]
85. Index - FHIR v4.0.1. HL7 International. URL: https://www.hl7.org/fhir/ [accessed 2020-07-07]

86. Trotta D, Garengo P. Industry 4.0 key research topics: A bibliometric review. 2018 Presented at: 2018 7th International Conference on Industrial Technology and Management (ICITM); March 7-9, 2018; Oxford, UK p. 113-117 URL: https://ieeexplore.ieee.org/document/8333930 [doi: 10.1109/icitm.2018.8333930]
87. Gensini G, Alderighi C, Rasoini R, Mazzanti M, Casolo G. Value of telemonitoring and telemedicine in heart failure management. Card Fail Rev 2017 Nov;3(2):116-121 [FREE Full text] [doi: 10.15420/cfr.2017:6:2] [Medline: 29387464]
88. Li KF. Smart home technology for telemedicine and emergency management. J Ambient Intell Humaniz Comput 2013;4(5):535-546 [FREE Full text] [doi: 10.1007/s12652-012-0129-8] [Medline: 32218875]
89. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
90. Almazan VKB, Mahipus FIB, Santos JRM, Samonte MJC. CAHM: companion animal health monitoring system. In: IC4E 2020: Proceedings of the 2020 11th International Conference on E-Education, E-Business, E-Management, and E-Learning. 2020 Presented at: 2020 11th International Conference on E-Education, E-Business, E-Management, and E-Learning; January 2020; Osaka, Japan p. 417-421. [doi: 10.1145/3377571.3377641]
91. McNicholas J, Gilbey A, Rennie A, Ahmedzai S, Dono JA, Ormerod E. Pet ownership and human health: a brief review of evidence and issues. BMJ 2005 Nov 26;331(7527):1252-1254 [FREE Full text] [doi: 10.1136/bmj.331.7527.1252] [Medline: 16308387]
92. Pastorinho MR, Sousa ACA. Pets as sentinels of human exposure to neurotoxic metals. In: Pets as Sentinels, Forecasters and Promoters of Human Health. Cham, Switzerland: Springer International Publishing; 2019.
93. Dodziuk H. Applications of 3D printing in healthcare. Kardiochir Torakochirurgia Pol 2016 Sep;13(3):283-293 [FREE Full text] [doi: 10.5114/kitp.2016.62625] [Medline: 27785150]
94. Dallas CE, Bell WC. Prediction modeling to determine the adequacy of medical response to urban nuclear attack. Disaster Med Public Health Prep 2007 Nov;1(2):80-89. [doi: 10.1097/DMP.0b013e318159a9e3] [Medline: 18388634]
95. Zhang H, Duan H, Zuo J, Song M, Zhang Y, Yang B, et al. Characterization of post-disaster environmental management for Hazardous Materials Incidents: Lessons learnt from the Tianjin warehouse explosion, China. J Environ Manage 2017 Sep 01;199:21-30. [doi: 10.1016/j.jenvman.2017.05.021] [Medline: 28525808]
96. Ullah Z, Al-Turjman F, Mostarda L, Gagliardi R. Applications of artificial intelligence and machine learning in smart cities. Comput Commun 2020 Mar 15;154:313-323 [FREE Full text] [doi: 10.1016/j.comcom.2020.02.069]
97. Foth M, Forlano L, Satchell C, Gibbs M, Donath J. From Social Butterfly to Engaged Citizen: Urban Informatics, Social Media, Ubiquitous Computing, and Mobile Technology to Support Citizen Engagement. Cambridge, Massachusetts: The MIT Press; 2011.
98. Bolívar MPR, Muñoz LA. E-Participation in Smart Cities: Technologies and Models of Governance for Citizen Engagement. Switzerland: Springer International Publishing; Aug 01, 2018.
99. Mukherjee S. Emerging frontiers in smart environment and healthcare – A vision. Inf Syst Front 2019 Dec 10;22(1):23-27 [FREE Full text] [doi: 10.1007/s10796-019-09965-3]
100. Lu Y. Industry 4.0: A survey on technologies, applications and open research issues. J Ind Inf Integr 2017 Jun;6:1-10. [doi: 10.1016/j.jii.2017.04.005]
101. Andrachuk M, Marschke M, Hings C, Armitage D. Smartphone technologies supporting community-based environmental monitoring and implementation: a systematic scoping review. Biol Conserv 2019 Sep;237:430-442. [doi: 10.1016/j.biocon.2019.07.026]
102. Neo JPS, Tan BH. The use of animals as a surveillance tool for monitoring environmental health hazards, human health hazards and bioterrorism. Vet Microbiol 2017 May;203:40-48 [FREE Full text] [doi: 10.1016/j.vetmic.2017.02.007] [Medline: 28619165]
103. Stieb DM, Huang A, Hocking R, Crouse DL, Osornio-Vargas AR, Villeneuve PJ. Using maps to communicate environmental exposures and health risks: Review and best-practice recommendations. Environ Res 2019 Sep;176:108518 [FREE Full text] [doi: 10.1016/j.envres.2019.05.049] [Medline: 31202044]
104. Majumder S, Aghayi E, Noferesti M, Memarzadeh-Tehran H, Mondal T, Pang Z, et al. Smart homes for elderly healthcare-recent advances and research challenges. Sensors (Basel) 2017 Oct 31;17(11):2496 [FREE Full text] [doi: 10.3390/s17112496] [Medline: 29088123]
105. Abdolkhani R, Gray K, Borda A, DeSouza R. Patient-generated health data management and quality challenges in remote patient monitoring. JAMIA Open 2019 Dec 20;2(4):471-478 [FREE Full text] [doi: 10.1093/jamiaopen/ooz036] [Medline: 32025644]
106. Beukenhorst AL, Howells K, Cook L, McBeth J, O'Neill TW, Parkes MJ, et al. Engagement and participant experiences with consumer smartwatches for health research: Longitudinal, observational feasibility study. JMIR Mhealth Uhealth 2020 Jan 29;8(1):e14368 [FREE Full text] [doi: 10.2196/14368] [Medline: 32012078]
107. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health 2020 Aug;2(8):e435-e440 [FREE Full text] [doi: 10.1016/S2589-7500(20)30142-4] [Medline: 32835201]

108. Leroy EM, Gouilh MA, Brugère-Picoux J. The risk of SARS-CoV-2 transmission to pets and other wild and domestic animals strongly mandates a one-health strategy to control the COVID-19 pandemic. One Health 2020 Dec;10:100133 [FREE Full text] [doi: 10.1016/j.onehlt.2020.100133] [Medline: 32363229]
109. Gortázar C, de la Fuente J. COVID-19 is likely to impact animal health. Prev Vet Med 2020 Jul;180:105030 [FREE Full text] [doi: 10.1016/j.prevetmed.2020.105030] [Medline: 32447153]
110. Torous J, Myrick KJ, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: Using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
111. Global strategy on digital health 2020–2025. World Health Organization. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf?sfvrsn=f112ede5_38 [accessed 2021-01-27]
112. Rüegg SR, Nielsen LR, Buttigieg SC, Santa M, Aragrande M, Canali M, et al. A systems approach to evaluate One Health initiatives. Front Vet Sci 2018 Mar 09;5:23 [FREE Full text] [doi: 10.3389/fvets.2018.00023] [Medline: 29594154]
113. Digital animalities. Social Sciences and Humanities Research Council of Canada. URL: https://digitalanimalities.org [accessed 2021-01-27]
114. Agrawal A. Environmentality: Technologies of Government and the Making of Subjects. Durham, North Carolina: Duke University Press; 2005.
115. Aydin C, Woge MG, Verbeek PP. Technological environmentality: Conceptualizing technology as a mediating milieu. Philosophy & Technology 2018 Apr 10;32(2):321-338 [FREE Full text] [doi: 10.1007/s13347-018-0309-3]
116. Mancini C. Animal-computer interaction: a manifesto. interactions 2011 Jul 01;18(4):69-73. [doi: 10.1145/1978822.1978836]
117. Barcelo D. An environmental and health perspective for COVID-19 outbreak: Meteorology and air quality influence, sewage epidemiology indicator, hospitals disinfection, drug therapies and recommendations. J Environ Chem Eng 2020 Aug;8(4):104006 [FREE Full text] [doi: 10.1016/j.jece.2020.104006] [Medline: 32373461]
118. Tamburis O. Bridging the gap between process mining and DES modeling in the healthcare domain. 2019 Presented at: 7th IEEE International Conference on E-Health and Bioengineering – E-HB 2019; 21-23/11/2019; Iasi, Romania.

Abbreviations

BMHI: biomedical and health informatics
EFMI: European Federation for Medical Informatics
FAIRER: findable, accessible, interoperable, reusable, ethical, and reversible
ODH: One Digital Health
WHO: World Health Organization

Edited by R Kukafka; submitted 08.07.20; peer-reviewed by A Ren, T Yano; comments to author 29.07.20; revised version received 09.08.20; accepted 24.01.21; published 05.02.21.

Please cite as:
Benis A, Tamburis O, Chronaki C, Moen A
One Digital Health: A Unified Framework for Future Health Ecosystems
J Med Internet Res 2021;23(2):e22189
URL: http://www.jmir.org/2021/2/e22189/
doi:10.2196/22189
PMID:33492240

Viewpoint

The Need for Ethnoracial Equity in Artificial Intelligence for Diabetes Management: Review and Recommendations

Quynh Pham[1,2], PhD; Anissa Gamble[1], MSc; Jason Hearn[1,3], MHSc; Joseph A Cafazzo[1,2,4], PEng, PhD

[1]Centre for Global eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

[2]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[3]Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada

[4]Institute of Biomedical Engineering, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Quynh Pham, PhD
Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto
Health Sciences Building
155 College Street
Toronto, ON, M5T 1P8
Canada
Phone: 1 4163404800 ext 4765
Email: q.pham@uhn.ca

Abstract

There is clear evidence to suggest that diabetes does not affect all populations equally. Among adults living with diabetes, those from ethnoracial minority communities—foreign-born, immigrant, refugee, and culturally marginalized—are at increased risk of poor health outcomes. Artificial intelligence (AI) is actively being researched as a means of improving diabetes management and care; however, several factors may predispose AI to ethnoracial bias. To better understand whether diabetes AI interventions are being designed in an ethnoracially equitable manner, we conducted a secondary analysis of 141 articles included in a 2018 review by Contreras and Vehi entitled "*Artificial Intelligence for Diabetes Management and Decision Support: Literature Review*." Two members of our research team independently reviewed each article and selected those reporting ethnoracial data for further analysis. Only 10 articles (7.1%) were ultimately selected for secondary analysis in our case study. Of the 131 excluded articles, 118 (90.1%) failed to mention participants' ethnic or racial backgrounds. The included articles reported ethnoracial data under various categories, including race (n=6), ethnicity (n=2), race/ethnicity (n=3), and percentage of Caucasian participants (n=1). Among articles specifically reporting race, the average distribution was 69.5% White, 17.1% Black, and 3.7% Asian. Only 2 articles reported inclusion of Native American participants. Given the clear ethnic and racial differences in diabetes biomarkers, prevalence, and outcomes, the inclusion of ethnoracial training data is likely to improve the accuracy of predictive models. Such considerations are imperative in AI-based tools, which are predisposed to negative biases due to their black-box nature and proneness to distributional shift. Based on our findings, we propose a short questionnaire to assess ethnoracial equity in research describing AI-based diabetes interventions. At this unprecedented time in history, AI can either mitigate or exacerbate disparities in health care. Future accounts of the infancy of diabetes AI must reflect our early and decisive action to confront ethnoracial inequities before they are coded into our systems and perpetuate the very biases we aim to eliminate. If we take deliberate and meaningful steps now toward training our algorithms to be ethnoracially inclusive, we can architect innovations in diabetes care that are bound by the diverse fabric of our society.

(J Med Internet Res 2021;23(2):e22320) doi:10.2196/22320

KEYWORDS

diabetes; artificial intelligence; digital health; ethnoracial equity; ethnicity; race

Introduction

There is clear evidence to suggest that diabetes does not affect all populations equally [1]. Among adults living with diabetes, those from ethnoracial minority communities—foreign-born, immigrant, refugee, and culturally marginalized [2]—are at increased risk of poor health outcomes [3-6]. Numerous studies have reported ethnoracial differences in glycemic control [7,8], diabetes prevalence [9], risk of diabetes complications [10], and diabetes-related mortality [11]. Data from the Centers for Disease Control indicate that non-Hispanic Black people are 2.3 times more likely to die from diabetes than their non-Hispanic White counterparts [12]. Similarly, young people living with diabetes from Black or Hispanic backgrounds are at increased risk of poor long-term glycemic control when compared to White youth [13]. The social determinants of health describe the social, economic, and physical conditions in which people are "born, live, learn, work, play, worship, and age," as well as the impact that such environments have on health outcomes [14]. As a result of the well-accepted contribution of the social determinants toward diabetes outcomes [15], we know that ethnoracial minority populations are also more likely to experience socioeconomic adversity and subsequent challenges with diabetes management and access to care [16]. This association likely follows from the increased prevalence of various diabetes risk factors (eg, low birth weight, physical inactivity, obesity, smoking) in individuals of low socioeconomic status (SES) [16-18]. In Canada, where 21% of the population are foreign-born and live in the nation's largest urban centers [2], people with a household income less than Can $15,000 (US $11,745) are 4 times more likely than those with a household income greater than Can $80,000 (US $62,635) to be diagnosed with type 2 diabetes (T2D) [19]. For people living with type 1 diabetes (T1D), low SES has been associated with an increased risk of poor glycemic control [20], as well as higher levels of mortality and morbidity [21].

Innovative technologies are actively being researched and developed to mitigate the burden of diabetes on patients and the health care system. Among the potential solutions, artificial intelligence (AI) appears to be well suited for diabetes management given that this chronic condition has long been guided by quantitative data collected by patients, their devices, and their care providers [22]. These data can be computationally complex for patients to make sense of on their own to inform their diabetes management [23]. AI, or the ability for "computers to think like humans" [24], has revolutionized many consumer technologies (eg, facial recognition, fraud detection, self-driving vehicles) and is now gaining momentum in the health care field. AI technologies are being developed for various areas of medicine such as medical imaging analysis [25-27], prognostication [28-30], and clinical decision support [31-33]. In diabetes care, AI is being applied for blood glucose prediction and control [34,35], identification of adverse events [36,37], lifestyle support [38,39], and predicting diabetes risk [40,41].

Despite the potential applications of AI in diabetes care, several factors may predispose these technologies to ethnoracial bias. The effectiveness of an AI algorithm is largely dependent on the quality of training data, as well as how accurately training data represent the population that will ultimately be affected by the algorithm [42]. As health data have traditionally been collected on predominantly White populations [43] or have simply omitted relevant ethnoracial information [44], algorithms trained on such data are at risk of ignoring race and ethnicity. Such ethnoracial disparities have long been present in clinical decision support tools, with various algorithms being arbitrarily corrected for race with little or no scientific justification [45]. These algorithms are widely used to inform important clinical actions such as specialist referrals [46,47] and assess candidacy for particular interventions [48,49]. In AI, where the effects of biases may be dramatic and difficult to identify [50], careless incorporation of ethnoracial data may perpetuate health inequities for those communities in most need. The alarming potential for clinical decision support tools to be algorithmically biased in favor of advantaged populations demands careful evaluation to promote their ethnoracial inclusivity. We believe that to optimize equitability, AI research should (1) establish a training population that is representative of the general population, (2) report the ethnoracial distribution of the training set, and (3) discuss potential ethnoracial limitations of the training data. To our knowledge, these simple tenets are not being met in existing diabetes AI research.

As a digital health research group preparing to build AI-based diabetes management tools [51], we want to address the challenges of promoting equity in AI and derive recommendations that can inform our work and the field at large. In an effort to better understand whether diabetes AI interventions are being designed in an ethnoracially equitable manner, we conducted a rapid case study whereby we assessed articles curated in an existing literature review of AI-based diabetes management tools. Our objectives were to (1) review ethnoracial considerations reported in past articles on AI-based diabetes support tools and (2) propose a strategy to promote ethnoracial equity in such tools in the future. This viewpoint serves to document the findings from our case study and the recommendations proposed by our group to advance ethnoracial equity in diabetes AI.

Case Study

Methods

We conducted a secondary analysis of 141 articles included in a high-impact literature review published in the Journal of Medical Internet Research in 2018 by Contreras and Vehi entitled "*Artificial Intelligence for Diabetes Management and Decision Support: Literature Review*" [52]. The selected review included articles describing AI technologies for diabetes management and decision support, as well as their associated challenges. We chose this review over comparable syntheses of the literature based on the short time since publication, the breadth of diabetes AI interventions included for review, and the impact that the review has had on informing the diabetes AI field.

Two members of our research team independently reviewed each of the 141 articles and selected those reporting ethnoracial data for further analysis. Articles were selected for analysis if they included an explicit description of participants' ethnic

background, racial background, or both. Articles were excluded if they were review papers, selected participants from a single ethnoracial group, or were inaccessible by the research team. The following criteria were charted for each of the selected studies: article type, diabetes type, ethnicity distribution, race distribution, number of participants, and source of data (ie, electronic medical record, electronic health record).

Results

In screening the 141 articles included in the Contreras and Vehi review, only 10 (7.1%) were ultimately selected for secondary analysis in our case study [53-62]. Of the 131 excluded articles, 118 (90.1%) failed to mention participants' ethnic or racial backgrounds. The remaining articles were excluded because they were review papers (n=5), selected participants from a single ethnoracial group (n=3), or were inaccessible by our research team (n=5).

The 10 articles selected for detailed analysis are summarized in Multimedia Appendices 1-3. Most articles were T2D-focused (n=8), with the remaining articles focused on T1D (n=1) and gestational diabetes (n=1). The main report types were retrospective analyses of data pulled from electronic medical records (n=5) or generated through randomized control trials (n=2). The reviewed articles reported ethnoracial data under various categories, including *race* (n=6), *ethnicity* (n=2), *race/ethnicity* (n=3), and *percentage of Caucasian participants* (n=1). Race was typically distributed between White (or Caucasian), Black (or African American), Asian, American Indian, and Alaska Native. Ethnicity was generally reported as Hispanic and non-Hispanic. Among articles specifically reporting race, the average distribution was 69.5% White, 17.1% Black, and 3.7% Asian (Multimedia Appendix 1). The 2 articles that specifically included ethnicity reported 7.2% and 21.3% Hispanic patients (Multimedia Appendix 2) [53,61]. The average distribution in articles that merged race and ethnicity was 55.4% non-Hispanic White, 8.1% non-Hispanic Black, 19.9% Hispanic, and 8.3% Asian (Multimedia Appendix 3). Only 2 articles reported inclusion of Native American participants [59,61]. The sole non-American study was performed in the Netherlands and included 97.7% Caucasian participants [60].

Several of the selected studies also included specific discussion of ethnoracial themes. Rohan et al stated that their research was limited by the homogeneity of their study population and that the generalizability of their findings should be further investigated [54]. Two more studies acknowledged that their study populations were mainly White [58,60], with one stating that their predominantly White and female demographic was "not uncommon in behavioral weight loss studies" [58]. Valdez et al intentionally oversampled racial and ethnic minorities and identified very few ethnoracial differences in health information communication patterns [61]. McCoy et al noted that race/ethnicity did not contribute to their predictions of glycemic trajectory and proposed that ethnoracial disparities in glycemic control may reflect differences in access to health care and medications [57].

Discussion

Ethnoracial Inequities in Diabetes AI

Diabetes AI programs are intended to improve diabetes-related health outcomes, experience, and expenditure [63,64]. However, it is unclear whether such systems benefit all populations equally. In our informal case study of 141 articles related to AI-based diabetes tools, we identified only 10 articles that specifically reported the ethnic or racial distribution of their studied patient population. We believe that this paucity of ethnoracial data in the reviewed articles significantly limits the effectiveness of the associated AI technologies. Several examples of such ethnoracial bias in clinical algorithms have been previously reported in the literature [42,45]. The long-used Framingham risk factors, which were modelled using a largely non-Hispanic White population, have recently been shown to inadequately capture risk in certain minority groups [65]. The STONE score to predict the likelihood of kidney stones in patients with flank pain equates Black race with lower risk [66]; however, an external validation study found no significant association between non-Black race and increased risk of developing kidney stones [67]. The Vaginal Birth after Cesarean (VBAC) algorithm predicts a lower likelihood of successful vaginal delivery in African American and Hispanic mothers having previously undergone cesarean section [68], while ignoring other factors (eg, private insurance status, marital status) that have been significantly associated with VBAC success [49]. A recent AI-based tool for classifying images of skin cancer was reported to perform similarly to trained experts [69]; however, the training images were predominantly of light-skinned individuals, and performance was not assessed on those with darker skin [70]. These examples highlight the importance of effective ethnoracial considerations in the development of clinical decision support tools.

Despite the promise of AI, several factors predispose AI algorithms to negative biases. One limitation of AI models is the so-called *distributional shift*, where erroneous predictions result from a mismatch between the training population and the population on which the model is used. Such a mismatch can result from "bias in the training set, change over time, or use of the system in a different population" [50]. Essentially, the robustness of AI algorithms is dependent upon the degree to which the training population represents the target population [71]. In addition to the distributional shift phenomenon, the complexity and black-box nature of AI algorithms often obfuscates underlying errors or biases, specifically when compared to simpler rule-based systems [50]. The detection of such biases in AI algorithms often requires careful consideration of model behavior in response to changing inputs [72]. In the case of ethnoracial data, the omission of such information could result in a distributional shift based on ethnicity, race, or both in resultant models, which may be difficult for researchers to identify at the time of development.

Given the clear ethnic and racial differences in diabetes biomarkers, prevalence, and outcomes [7-10,12,73], the inclusion of ethnoracial data is also likely to improve the accuracy of predictive models. The predictive value of race and

ethnicity is well-documented in the literature, where they have been shown to independently predict health decline for adults living with diabetes [74,75]. The impact of specific risk factors for T2D have even been shown to vary for both sex and race, with the most predictive factors being waist circumference in Black men, 2-hour glucose from an oral glucose tolerance test in Black women, and fasting glucose in both White men and White women [76]. As a result of these associations between diabetes outcomes and ethnoracial information, the consideration of ethnoracial data is likely to enhance both the accuracy and generalizability of resultant AI-based diabetes tools.

In those articles we reviewed that did include ethnoracial information, there was very little standardization in terms of how these data were reported (eg, *race*, *race and ethnicity*, *race/ethnicity*). Race distinguishes individuals based on ancestry and combinations of physical characteristics, whereas ethnicity focuses on behavior and culture in addition to physical features [77]. Inconsistent reporting of ethnic and racial information hinders the ability to perform meta-analyses across multiple data sets and may limit ethnoracial equity in future AI applications [78]. In their writings on eliminating health disparities, Fremont and Lurie state that data pertaining to race and ethnicity are collected by a variety of sources, but "the utility of these data is constrained by ongoing problems with reliability, completeness, and lack of comparability across data sources" [79]. Though differences in the reporting of ethnoracial data are expected across jurisdictions, we propose that authors attempt to report such data in a manner that is easily comparable to locally available data. For example, the US census reports race and ethnicity separately, with ethnicity being used to determine whether an individual is of "Hispanic origin or not" and race being categorized as "White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, and other" [80]. Kiran et al recently assessed Canadian patient perspectives on routinely being asked about their race and ethnicity through a sociodemographic questionnaire [81]. They found that patients were not uncomfortable disclosing their race and ethnicity and intuitively understood how the data could be helpful for their health care providers. Their work has subsequently informed the collection of race-based data during the COVID-19 pandemic [82]. These are just two examples of standards that will allow for comparison with locally available data and in turn enable the assessment of ethnoracial generalizability and cultural competence in diabetes AI algorithms.

In considering the average race distribution of the reviewed studies, the proportions for White (69.5%) and Asian race (3.7%) were slightly lower than those values reported in recent US census data (76.3% and 5.9%, respectively). The opposite was true for the Black race, which accounted for 17.1% of study participants but only 13.4% of US census participants. In those studies reporting race and ethnicity as a combined variable, the average proportion of non-Hispanic whites (55.4%) was slightly lower than the census value of 60.1% [83]. These findings likely follow from the high prevalence of diabetes in the non-Hispanic Black population, specifically when compared to the non-Hispanic White and Asian populations [84]. One particularly worrisome finding was that data from Native American participants were reported in just 2 studies [59,61], despite making up an estimated 1.3% of the American population [83] and being the ethnoracial group with the highest age-adjusted prevalence of diagnosed diabetes [9]. Poor Indigenous representation in health and governmental data sets has been previously reported in the literature [85,86]. In Canada, where Indigenous peoples account for 4.9% of the population [87] and are disproportionately affected by diabetes [88], failure to include Indigenous data when training diabetes AI models could propagate existing issues of health inequity and structural racism in this population [89,90].

A Simple Screening Tool to Assess Ethnoracial Equity in Diabetes AI

Detailed guidelines currently exist for the development of trustworthy and human-centric AI technologies [91]. However, we believe there is a need for simple tools to screen the ethnic and racial generalizability of AI in health care. Based on the findings from our case study, we have developed a short screening tool that researchers and clinicians may use to assess ethnoracial equity in research describing AI-based diabetes interventions. The rationale and structure of this tool borrows from the Jadad scale [92], which was conceived by the founder of our research group over two decades ago and is widely used to assess the methodological quality of a clinical trial [93,94]. We propose the following set of five questions to consider the ethnoracial relevance of diabetes AI:

1. Did the research explicitly describe the disease under study (eg, T1D, T2D, both)? (1a) Did the research describe ethnoracial differences in disease prevalence, biomarkers, and outcomes?
2. Did the research clearly describe the sources of data used in the training data set (eg, electronic medical record, administrative data repository, research registry)? (2a) Did the research describe ethnoracial limitations in the sources of data?
3. Did the research explicitly report the ethnic and racial backgrounds of individuals in the training data set? (3a) Are ethnic and racial backgrounds reported in a manner that is easily comparable to local census data?
4. Do the ethnic and racial distributions in the training data set accurately represent the population on which the algorithm will be used? (4a) Did the research articulate limitations of the ethnic and racial distributions in the training data set?
5. Did the research describe strategies to mitigate ethnoracial bias in the algorithm?

Although we feel that our proposed tool will be helpful in assessing clinical AI algorithms generally, it will be particularly important in the development of diabetes AI. We believe that these innovations will fail to serve the diabetes community if they are not trained on ethnoracially diverse data. As AI-based systems become integrated into important aspects of diabetes management, such ethnoracial inequities in model development could ultimately be dangerous for minority groups whose biomarkers and outcomes may differ from the general population. In the Contreras and Vehi review, most studies focused on T2D self-management, clinical decision support,

and prediction tools. Each of these dimensions of diabetes care can be affected by ethnoracial factors. For example, adherence to T2D medications to achieve euglycemia is demonstrably driven by cultural beliefs, values, social factors, religion, health literacy, and language barriers [95,96]. Similar issues are likely to follow in the T1D space, where AI algorithms are currently focused on automated insulin delivery systems but will likely shift toward the above dimensions in the near future [63,97,98].

Addressing ethnoracial bias in diabetes AI has been made even more critical by the coronavirus disease 2019 (COVID-19) pandemic [99]. There is growing evidence to support a "bidirectional relationship between COVID-19 and diabetes" [100]. Research suggests that diabetes is a risk factor for rapid progression and poor prognosis of COVID-19 [101,102]. New-onset diabetes is also being reported in previously healthy individuals diagnosed with COVID-19 [103-105], which may reflect coronavirus-inflicted damage to insulin-producing cells [106,107]. We are concerned by these findings from a health equity lens, given that COVID-19 has been found to disproportionately affect ethnoracial minorities. The Centers for Disease Control and Prevention have already determined that individuals from Black and American Indian or Alaska Native communities have a rate of hospitalization or death from COVID-19 that is 5 times greater than that of their White counterparts [108]. It stands to reason that the increased prevalence of both COVID-19 and diabetes in ethnoracial minority groups and the relationship between these two conditions require ethnoracial considerations in all aspects of diabetes care.

At this unprecedented time in history, AI can either mitigate or exacerbate disparities in health care. Future accounts of the infancy of diabetes AI must reflect our early and decisive action to confront ethnoracial inequities before they are coded into our systems and perpetuate the very biases we aim to eliminate [45]. If we take deliberate and meaningful steps now toward training our algorithms to be ethnoracially inclusive, we can architect innovations in diabetes care that are bound by the diverse fabric of our society.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution of articles specifically reporting race.
[DOCX File , 14 KB - jmir_v23i2e22320_app1.docx]

Multimedia Appendix 2
Distribution of articles specifically reporting ethnicity.
[DOCX File , 13 KB - jmir_v23i2e22320_app2.docx]

Multimedia Appendix 3
Distribution for articles reporting race and ethnicity as a merged variable.
[DOCX File , 14 KB - jmir_v23i2e22320_app3.docx]

References

1. Piccolo RS, Subramanian S, Pearce N, Florez JC, McKinlay JB. Relative Contributions of Socioeconomic, Local Environmental, Psychosocial, Lifestyle/Behavioral, Biophysiological, and Ancestral Factors to Racial/Ethnic Disparities in Type 2 Diabetes. Diabetes Care 2016 Jul 21;39(7):1208-1217 [FREE Full text] [doi: 10.2337/dc15-2255] [Medline: 27330127]
2. Immigration and Ethnocultural Diversity in Canada. 2013. URL: https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-010-x/99-010-x2011001-eng.cfm [accessed 2021-01-16]
3. Williams DR, Priest N, Anderson NB. Understanding associations among race, socioeconomic status, and health: Patterns and prospects. Health Psychol 2016 May;35(4):407-411 [FREE Full text] [doi: 10.1037/hea0000242] [Medline: 27018733]
4. Isaranuwatchai W, Fazli GS, Bierman AS, Lipscombe LL, Mitsakakis N, Shah BR, et al. Universal Drug Coverage and Socioeconomic Disparities in Health Care Costs Among Persons With Diabetes. Diabetes Care 2020 Oct;43(9):2098-2105. [doi: 10.2337/dc19-1536] [Medline: 32641377]
5. Agardh E, Allebeck P, Hallqvist J, Moradi T, Sidorchuk A. Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. Int J Epidemiol 2011 Jul 19;40(3):804-818. [doi: 10.1093/ije/dyr029] [Medline: 21335614]
6. Bird Y, Lemstra M, Rogers M, Moraros J. The relationship between socioeconomic status/income and prevalence of diabetes and associated conditions: A cross-sectional population-based study in Saskatchewan, Canada. Int J Equity Health 2015 Oct 12;14(1):93 [FREE Full text] [doi: 10.1186/s12939-015-0237-0] [Medline: 26458543]
7. Carson AP, Muntner P, Selvin E, Carnethon MR, Li X, Gross MD, et al. Do glycemic marker levels vary by race? Differing results from a cross-sectional analysis of individuals with and without diagnosed diabetes. BMJ Open Diabetes Res Care 2016 Jun 10;4(1):e000213 [FREE Full text] [doi: 10.1136/bmjdrc-2016-000213] [Medline: 27335652]

8. Cavagnolli G, Pimentel AL, Freitas PAC, Gross JL, Camargo JL. Effect of ethnicity on HbA1c levels in individuals without diabetes: Systematic review and meta-analysis. PLoS One 2017 Feb 13;12(2):e0171315 [FREE Full text] [doi: 10.1371/journal.pone.0171315] [Medline: 28192447]
9. Spanakis EK, Golden SH. Race/ethnic difference in diabetes and diabetic complications. Curr Diab Rep 2013 Dec 15;13(6):814-823 [FREE Full text] [doi: 10.1007/s11892-013-0421-9] [Medline: 24037313]
10. Lanting LC, Joung IM, Mackenbach JP, Lamberts SW, Bootsma AH. Ethnic differences in mortality, end-stage complications, and quality of care among diabetic patients: a review. Diabetes Care 2005 Oct 25;28(9):2280-2288. [doi: 10.2337/diacare.28.9.2280] [Medline: 16123507]
11. McBean AM, Li S, Gilbertson DT, Collins AJ. Differences in diabetes prevalence, incidence, and mortality among the elderly of four racial/ethnic groups: whites, blacks, hispanics, and asians. Diabetes Care 2004 Oct 24;27(10):2317-2324. [doi: 10.2337/diacare.27.10.2317] [Medline: 15451894]
12. Golden S, Brown A, Cauley JA, Chin MH, Gary-Webb TL, Kim C, et al. Health disparities in endocrine disorders: biological, clinical, and nonclinical factors--an Endocrine Society scientific statement. J Clin Endocrinol Metab 2012 Oct;97(9):E1579-E1639 [FREE Full text] [doi: 10.1210/jc.2012-2043] [Medline: 22730516]
13. Kahkoska AR, Shay CM, Crandell J, Dabelea D, Imperatore G, Lawrence JM, et al. Association of Race and Ethnicity With Glycemic Control and Hemoglobin A Levels in Youth With Type 1 Diabetes. JAMA Netw Open 2018 Oct 07;1(5):e181851 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.1851] [Medline: 30370425]
14. Healthy People 2020: An opportunity to address societal determinants of health in the United States. In: Secretary's Advisory Committee on Health Promotion and Disease Prevention Objectives for 2020. Washington, DC: US Department of Health and Human Services; 2010.
15. Walker RJ, Strom Williams J, Egede LE. Influence of Race, Ethnicity and Social Determinants of Health on Diabetes Outcomes. Am J Med Sci 2016 May;351(4):366-373 [FREE Full text] [doi: 10.1016/j.amjms.2016.01.008] [Medline: 27079342]
16. Caruso R, Magon A, Baroni I, Dellafiore F, Arrigoni C, Pittella F, et al. Health literacy in type 2 diabetes patients: a systematic review of systematic reviews. Acta Diabetol 2018 Jan 11;55(1):1-12. [doi: 10.1007/s00592-017-1071-1] [Medline: 29129000]
17. Connolly V, Unwin N, Sherriff P, Bilous R, Kelly W. Diabetes prevalence and socioeconomic status: a population based study showing increased prevalence of type 2 diabetes mellitus in deprived areas. J Epidemiol Community Health 2000 Mar;54(3):173-177 [FREE Full text] [doi: 10.1136/jech.54.3.173] [Medline: 10746110]
18. Elbein SC. Genetics factors contributing to type 2 diabetes across ethnicities. J Diabetes Sci Technol 2009 Jul 01;3(4):685-689 [FREE Full text] [doi: 10.1177/193229680900300412] [Medline: 20144314]
19. Dinca-Panaitescu S, Dinca-Panaitescu M, Bryant T, Daiski I, Pilkington B, Raphael D. Diabetes prevalence and income: Results of the Canadian Community Health Survey. Health Policy 2011 Mar;99(2):116-123. [doi: 10.1016/j.healthpol.2010.07.018] [Medline: 20724018]
20. Gallegos-Macias A, Macias SR, Kaufman E, Skipper B, Kalishman N. Relationship between glycemic control, ethnicity and socioeconomic status in Hispanic and white non-Hispanic youths with type 1 diabetes mellitus. Pediatr Diabetes 2003 Mar;4(1):19-23. [doi: 10.1034/j.1399-5448.2003.00020.x] [Medline: 14655519]
21. Scott A, Chambers D, Goyder E, O'Cathain A. Socioeconomic inequalities in mortality, morbidity and diabetes management for adults with type 1 diabetes: A systematic review. PLoS One 2017 May 10;12(5):e0177210 [FREE Full text] [doi: 10.1371/journal.pone.0177210] [Medline: 28489876]
22. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine Learning and Data Mining Methods in Diabetes Research. Comput Struct Biotechnol J 2017;15:104-116 [FREE Full text] [doi: 10.1016/j.csbj.2016.12.005] [Medline: 28138367]
23. Gunasekeran DV. Technology and chronic disease management. The Lancet Diabetes & Endocrinology 2018 Feb;6(2):91. [doi: 10.1016/s2213-8587(17)30441-2]
24. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business Horizons 2019 Jan;62(1):15-25. [doi: 10.1016/j.bushor.2018.08.004]
25. Tang A, Tam R, Cadrin-Chênevert A, Guest W, Chong J, Barfett J, Canadian Association of Radiologists (CAR) Artificial Intelligence Working Group. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. Can Assoc Radiol J 2018 May 01;69(2):120-135 [FREE Full text] [doi: 10.1016/j.carj.2018.02.002] [Medline: 29655580]
26. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp 2018 Oct 24;2(1):35 [FREE Full text] [doi: 10.1186/s41747-018-0061-6] [Medline: 30353365]
27. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. Radiology 2020 Aug 19;296(2):E65-E71 [FREE Full text] [doi: 10.1148/radiol.2020200905] [Medline: 32191588]
28. Jerez-Aragonés JM, Gómez-Ruiz JA, Ramos-Jiménez G, Muñoz-Pérez J, Alba-Conejo E. A combined neural network and decision trees model for prognosis of breast cancer relapse. Artif Intell Med 2003 Jan;27(1):45-63. [doi: 10.1016/s0933-3657(02)00086-6] [Medline: 12473391]

29. Liu X, Chen K, Wu T, Weidman D, Lure F, Li J. Use of multimodality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer's disease. Transl Res 2018 Apr;194:56-67 [FREE Full text] [doi: 10.1016/j.trsl.2018.01.001] [Medline: 29352978]
30. Hearn J, Ross HJ, Mueller B, Fan C, Crowdy E, Duhamel J, et al. Neural Networks for Prognostication of Patients With Heart Failure. Circ Heart Fail 2018 Aug;11(8):e005193. [doi: 10.1161/CIRCHEARTFAILURE.118.005193] [Medline: 30354561]
31. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLoS One 2017 Apr 6;12(4):e0174708 [FREE Full text] [doi: 10.1371/journal.pone.0174708] [Medline: 28384212]
32. Valdes G, Simone CB, Chen J, Lin A, Yom SS, Pattison AJ, et al. Clinical decision support of radiotherapy treatment planning: A data-driven machine learning strategy for patient-specific dosimetric decision making. Radiother Oncol 2017 Dec;125(3):392-397. [doi: 10.1016/j.radonc.2017.10.014] [Medline: 29162279]
33. Peiffer-Smadja N, Rawson T, Ahmad R, Buchard A, Georgiou P, Lescure F, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clin Microbiol Infect 2020 May;26(5):584-595. [doi: 10.1016/j.cmi.2019.09.009] [Medline: 31539636]
34. Plis K, Bunescu R, Marling C, Shubrook J, Schwartz F. A machine learning approach to predicting blood glucose levels for diabetes management. 2014 Presented at: Workshops at the Twenty-Eighth AAAI conference on artificial intelligence; July 27-31, 2014; Quebec City, QC.
35. Gyuk P, Vassányi I, Kósa I. Blood Glucose Level Prediction for Diabetics Based on Nutrition and Insulin Administration Logs Using Personalized Mathematical Models. J Healthc Eng 2019 Jan 10;2019:8605206-8605212 [FREE Full text] [doi: 10.1155/2019/8605206] [Medline: 30774850]
36. Lagani V, Chiarugi F, Manousos D, Verma V, Fursse J, Marias K, et al. Realization of a service for the long-term risk assessment of diabetes-related complications. J Diabetes Complications 2015 Jul;29(5):691-698. [doi: 10.1016/j.jdiacomp.2015.03.011] [Medline: 25953402]
37. Dagliati A, Marini S, Sacchi L, Cogni G, Teliti M, Tibollo V, et al. Machine Learning Methods to Predict Diabetes Complications. J Diabetes Sci Technol 2018 Mar 12;12(2):295-302 [FREE Full text] [doi: 10.1177/1932296817706375] [Medline: 28494618]
38. Fioravanti A, Fico G, Arredondo MT, Leuteritz JP. A mobile feedback system for integrated E-health platforms to improve self-care and compliance of diabetes mellitus patients. In: Conf Proc IEEE Eng Med Biol Soc. 2011 Presented at: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 30-September 3, 2011; Boston, MA p. 3550-3553. [doi: 10.1109/iembs.2011.6090591]
39. Caballero-Ruiz E, García-Sáez G, Rigla M, Villaplana M, Pons B, Hernando ME. A web-based clinical decision support system for gestational diabetes: Automatic diet prescription and detection of insulin needs. Int J Med Inform 2017 Jun;102:35-49. [doi: 10.1016/j.ijmedinf.2017.02.014] [Medline: 28495347]
40. Sowjanya K, Singhal A, Choudhary C. MobDBTest: A machine learning based system for predicting diabetes risk using mobile devices. 2015 Presented at: 2015 IEEE International Advance Computing Conference (IACC); 2015; Banglore p. 397-402. [doi: 10.1109/IADCC.2015.7154738]
41. Zou Q, Qu K, Luo Y, Yin D, Ju Y, Tang H. Predicting Diabetes Mellitus With Machine Learning Techniques. Front Genet 2018 Nov 6;9:515 [FREE Full text] [doi: 10.3389/fgene.2018.00515] [Medline: 30459809]
42. Parikh RB, Teeple S, Navathe AS. Addressing Bias in Artificial Intelligence in Health Care. JAMA 2019 Dec 22;322(24):2377 [FREE Full text] [doi: 10.1001/jama.2019.18058] [Medline: 31755905]
43. Geneviève LD, Martani A, Shaw D, Elger BS, Wangmo T. Structural racism in precision medicine: leaving no one behind. BMC Med Ethics 2020 Feb 19;21(1):17 [FREE Full text] [doi: 10.1186/s12910-020-0457-8] [Medline: 32075640]
44. Sultana K, Sheikh A. Most UK datasets of routinely collected health statistics fail to collect information on ethnicity and religion. J R Soc Med 2008 Oct;101(9):463-465 [FREE Full text] [doi: 10.1258/jrsm.2008.080007] [Medline: 18779248]
45. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight - Reconsidering the Use of Race Correction in Clinical Algorithms. N Engl J Med 2020 Aug 27;383(9):874-882 [FREE Full text] [doi: 10.1056/NEJMms2004740] [Medline: 32853499]
46. Peterson PN, Rumsfeld JS, Liang L, Albert NM, Hernandez AF, Peterson ED, American Heart Association Get With the Guidelines-Heart Failure Program. A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association get with the guidelines program. Circ Cardiovasc Qual Outcomes 2010 Jan;3(1):25-32. [doi: 10.1161/CIRCOUTCOMES.109.854877] [Medline: 20123668]
47. Eneanya ND, Yang W, Reese PP. Reconsidering the Consequences of Using Race to Estimate Kidney Function. JAMA 2019 Jul 09;322(2):113-114. [doi: 10.1001/jama.2019.5774] [Medline: 31169890]
48. Shahian DM, Jacobs JP, Badhwar V, Kurlansky PA, Furnary AP, Cleveland JC, et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 1-Background, Design Considerations, and Model Development. Ann Thorac Surg 2018 May;105(5):1411-1418. [doi: 10.1016/j.athoracsur.2018.03.002] [Medline: 29577925]
49. Landon M, Leindecker S, Spong CY, Hauth JC, Bloom S, Varner MW, National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. The MFMU Cesarean Registry: factors affecting the success of trial

of labor after previous cesarean delivery. Am J Obstet Gynecol 2005 Oct;193(3 Pt 2):1016-1023. [doi: 10.1016/j.ajog.2005.05.066] [Medline: 16157104]
50. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf 2019 Mar 12;28(3):231-237 [FREE Full text] [doi: 10.1136/bmjqs-2018-008370] [Medline: 30636200]
51. Centre for Global eHealth Innovation. URL: https://ehealthinnovation.org/ [accessed 2021-01-31]
52. Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res 2018 May 30;20(5):e10775 [FREE Full text] [doi: 10.2196/10775] [Medline: 29848472]
53. Nunes AP, Yang J, Radican L, Engel SS, Kurtyka K, Tunceli K, et al. Assessing occurrence of hypoglycemia and its severity from electronic health records of patients with type 2 diabetes mellitus. Diabetes Res Clin Pract 2016 Dec;121:192-203. [doi: 10.1016/j.diabres.2016.09.012] [Medline: 27744128]
54. Rohan J, Delamater A, Pendley JS, Dolan L, Reeves G, Drotar D. Identification of self-management patterns in pediatric type 1 diabetes using cluster analysis. Pediatr Diabetes 2011 Dec;12(7):611-618 [FREE Full text] [doi: 10.1111/j.1399-5448.2010.00752.x] [Medline: 21446925]
55. Anderson JP, Parikh JR, Shenfeld DK, Ivanov V, Marks C, Church BW, et al. Reverse Engineering and Evaluation of Prediction Models for Progression to Type 2 Diabetes: An Application of Machine Learning Using Electronic Health Records. J Diabetes Sci Technol 2015 Dec 20;10(1):6-18 [FREE Full text] [doi: 10.1177/1932296815620200] [Medline: 26685993]
56. DuBrava S, Mardekian J, Sadosky A, Bienen EJ, Parsons B, Hopps M, et al. Using Random Forest Models to Identify Correlates of a Diabetic Peripheral Neuropathy Diagnosis from Electronic Health Record Data. Pain Med 2017 Jan 01;18(1):107-115. [doi: 10.1093/pm/pnw096] [Medline: 27252307]
57. McCoy RG, Ngufor C, Van Houten HK, Caffo B, Shah ND. Trajectories of Glycemic Change in a National Cohort of Adults With Previously Controlled Type 2 Diabetes. Med Care 2017 Nov;55(11):956-964 [FREE Full text] [doi: 10.1097/MLR.0000000000000807] [Medline: 28922296]
58. Everett E, Kane B, Yoo A, Dobs A, Mathioudakis N. A Novel Approach for Fully Automated, Personalized Health Coaching for Adults with Prediabetes: Pilot Clinical Trial. J Med Internet Res 2018 Feb 27;20(2):e72 [FREE Full text] [doi: 10.2196/jmir.9723] [Medline: 29487046]
59. Hazlehurst BL, Lawrence JM, Donahoo WT, Sherwood NE, Kurtz SE, Xu S, et al. Automating assessment of lifestyle counseling in electronic health records. Am J Prev Med 2014 May;46(5):457-464 [FREE Full text] [doi: 10.1016/j.amepre.2014.01.001] [Medline: 24745635]
60. Cleveringa FG, Welsing PM, van den Donk M, Gorter KJ, Niessen LW, Rutten GE, et al. Cost-effectiveness of the diabetes care protocol, a multifaceted computerized decision support diabetes management intervention that reduces cardiovascular risk. Diabetes Care 2010 Mar 23;33(2):258-263 [FREE Full text] [doi: 10.2337/dc09-1232] [Medline: 19933991]
61. Valdez R, Guterbock TM, Fitzgibbon K, Williams IC, Wellbeloved-Stone CA, Bears JE, et al. From loquacious to reticent: understanding patient health information communication to guide consumer health IT design. J Am Med Inform Assoc 2017 Jul 01;24(4):680-696 [FREE Full text] [doi: 10.1093/jamia/ocw155] [Medline: 28069667]
62. Lo-Ciganic W, Donohue JM, Thorpe JM, Perera S, Thorpe CT, Marcum ZA, et al. Using machine learning to examine medication adherence thresholds and risk of hospitalization. Med Care 2015 Aug;53(8):720-728 [FREE Full text] [doi: 10.1097/MLR.0000000000000394] [Medline: 26147866]
63. Dankwa-Mullan I, Rivo M, Sepulveda M, Park Y, Snowdon J, Rhee K. Transforming Diabetes Care Through Artificial Intelligence: The Future Is Here. Popul Health Manag 2019 Jun;22(3):229-242 [FREE Full text] [doi: 10.1089/pop.2018.0129] [Medline: 30256722]
64. Singla R, Singla A, Gupta Y, Kalra S. Artificial Intelligence/Machine Learning in Diabetes Care. Indian J Endocrinol Metab 2019;23(4):495-497 [FREE Full text] [doi: 10.4103/ijem.IJEM_228_19] [Medline: 31741913]
65. Gijsberts CM, Groenewegen KA, Hoefer IE, Eijkemans MJC, Asselbergs FW, Anderson TJ, et al. Race/Ethnic Differences in the Associations of the Framingham Risk Factors with Carotid IMT and Cardiovascular Events. PLoS One 2015 Jul 2;10(7):e0132321 [FREE Full text] [doi: 10.1371/journal.pone.0132321] [Medline: 26134404]
66. Moore C, Bomann S, Daniels B, Luty S, Molinaro A, Singh D, et al. Derivation and validation of a clinical prediction rule for uncomplicated ureteral stone--the STONE score: retrospective and prospective observational cohort studies. BMJ 2014 Mar 26;348:g2191 [FREE Full text] [doi: 10.1136/bmj.g2191] [Medline: 24671981]
67. Wang RC, Rodriguez RM, Moghadassi M, Noble V, Bailitz J, Mallin M, et al. External Validation of the STONE Score, a Clinical Prediction Rule for Ureteral Stone: An Observational Multi-institutional Study. Ann Emerg Med 2016 May;67(4):423-432.e2 [FREE Full text] [doi: 10.1016/j.annemergmed.2015.08.019] [Medline: 26440490]
68. Grobman W, Lai Y, Landon MB, Spong CY, Leveno KJ, Rouse DJ, National Institute of Child Health and Human Development (NICHD) Maternal-Fetal Medicine Units Network (MFMU). Development of a nomogram for prediction of vaginal birth after cesarean delivery. Obstet Gynecol 2007 May;109(4):806-812. [doi: 10.1097/01.AOG.0000259312.36053.02] [Medline: 17400840]
69. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Feb 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]

70. Zou J, Schiebinger L. AI can be sexist and racist - it's time to make it fair. Nature 2018 Jul 18;559(7714):324-326. [doi: 10.1038/d41586-018-05707-8] [Medline: 30018439]
71. Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J, Mane D. Concrete Problems in AI Safety. arXiv. Preprint posted online on July 25, 2016.
72. Adler P, Falk C, Friedler SA, Nix T, Rybeck G, Scheidegger C, et al. Auditing black-box models for indirect influence. Knowl Inf Syst 2017 Oct 25;54(1):95-122. [doi: 10.1007/s10115-017-1116-3]
73. Brancati FL, Whelton PK, Kuller LH, Klag MJ. Diabetes mellitus, race, and socioeconomic status. A population-based study. Ann Epidemiol 1996 Jan;6(1):67-73. [doi: 10.1016/1047-2797(95)00095-x] [Medline: 8680628]
74. DeFronzo RA, Ferrannini E, Groop L, Henry RR, Herman WH, Holst JJ, et al. Type 2 diabetes mellitus. Nat Rev Dis Primers 2015 Jul 23;1:15019. [doi: 10.1038/nrdp.2015.19] [Medline: 27189025]
75. Nicklett EJ. Socioeconomic status and race/ethnicity independently predict health decline among older diabetics. BMC Public Health 2011 Oct 02;11(1):684 [FREE Full text] [doi: 10.1186/1471-2458-11-684] [Medline: 21888645]
76. Guo F, Garvey WT. Development of a Weighted Cardiometabolic Disease Staging (CMDS) System for the Prediction of Future Diabetes. J Clin Endocrinol Metab 2015 Oct;100(10):3871-3877 [FREE Full text] [doi: 10.1210/jc.2015-2691] [Medline: 26241327]
77. Edwards C, Fillingim RB, Keefe F. Race, ethnicity and pain. Pain 2001 Dec;94(2):133-137. [doi: 10.1016/S0304-3959(01)00408-0] [Medline: 11690726]
78. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. J Am Med Inform Assoc 2020 Dec 09;27(12):2011-2015 [FREE Full text] [doi: 10.1093/jamia/ocaa088] [Medline: 32594179]
79. National Research Council (US) Panel on DHHS Collection of Race and Ethnic Data, Ver Ploeg M, Perrin E. The Role of Racial and Ethnic Data Collection in Eliminating Disparities in Health Care. In: Eliminating Health Disparities: Measurement and Data Needs. Washington, DC: National Academies Press (US); 2004.
80. Jones NA, U.S. Census Bureau. Update on the U.S. Census Bureau's Race and Ethnic Research for the 2020 Census. URL: https://www.census.gov/content/dam/Census/newsroom/press-kits/2014/article_race_ethnic_research_2020census_jones.pdf [accessed 2021-01-16]
81. Kiran T, Sandhu P, Aratangy T, Devotta K, Lofters A, Pinto AD. Patient perspectives on routinely being asked about their race and ethnicity: Qualitative study in primary care. Can Fam Physician 2019 Aug;65(8):e363-e369 [FREE Full text] [Medline: 31413042]
82. Pinto AD, Hapsari A. Collecting data on race during the COVID-19 pandemic to identify inequities.: Upstream Lab URL: https://upstreamlab.files.wordpress.com/2020/04/upstream-lab-collecting-data-on-race-during-covid-brief-final-for-circulation-1.pdf [accessed 2021-01-16]
83. United States Census Bureau QuickFacts. URL: https://www.census.gov/quickfacts/fact/table/US/PST045219 [accessed 2021-01-16]
84. Cheng YJ, Kanaya AM, Araneta MRG, Saydah SH, Kahn HS, Gregg EW, et al. Prevalence of Diabetes by Race and Ethnicity in the United States, 2011-2016. JAMA 2019 Dec 24;322(24):2389-2398 [FREE Full text] [doi: 10.1001/jama.2019.19365] [Medline: 31860047]
85. Fremantle E, Zurynski YA, Mahajan D, D'Antoine H, Elliott EJ. Indigenous child health: urgent need for improved data to underpin better health outcomes. Med J Aust 2008 May 19;188(10):588-591. [doi: 10.5694/j.1326-5377.2008.tb01797.x] [Medline: 18484933]
86. Trevethan S. Strengthening the Availability of First Nations Data.: Indigenous Services Canada & The Assembly of First Nations; 2019. URL: https://www.afn.ca/wp-content/uploads/2019/05/NCR-11176060-v1-STRENGTHENING_THE_AVAILABILITY_OF_FIRST_NATIONS_DATA-MAR_25_2019-FINAL_E.pdf [accessed 2021-01-16]
87. Aboriginal peoples in Canada: Key results from the 2016 Census.: Government of Canada, Statistics Canada; 2017. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/171025/dq171025a-eng.htm [accessed 2021-01-31]
88. Dyck R, Osgood N, Lin TH, Gao A, Stang MR. Epidemiology of diabetes mellitus among First Nations and non-First Nations adults. CMAJ 2010 Mar 23;182(3):249-256 [FREE Full text] [doi: 10.1503/cmaj.090846] [Medline: 20083562]
89. Greenwood M, de Leeuw S, Lindsay N. Challenges in health equity for Indigenous peoples in Canada. Lancet 2018 Apr 28;391(10131):1645-1648. [doi: 10.1016/S0140-6736(18)30177-6] [Medline: 29483024]
90. Paradies Y. Colonisation, racism and indigenous health. J Pop Research 2016 Feb 18;33(1):83-96. [doi: 10.1007/s12546-016-9159-y]
91. Lemonne E. Ethics Guidelines for Trustworthy AI.: FUTURIUM - European Commission; 2018. URL: https://ec.europa.eu/futurium/en/ai-alliance-consultation [accessed 2021-01-31]
92. Jadad AR, Moore R, Carroll D, Jenkinson C, Reynolds DM, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996 Mar;17(1):1-12. [doi: 10.1016/0197-2456(95)00134-4] [Medline: 8721797]
93. Olivo S, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. Phys Ther 2008 Mar;88(2):156-175. [doi: 10.2522/ptj.20070147] [Medline: 18073267]

94. Halpern SH, Douglas MJ, editors. Appendix: Jadad Scale for Reporting Randomized Controlled Trials. In: Evidence-based Obstetric Anesthesia. Hoboken, NJ: Blackwell Publishing Ltd; Jun 2005:237-238.
95. Riddle MC. The underuse of insulin therapy in North America. Diabetes Metab Res Rev 2002 Sep;18 Suppl 3(S3):S42-S49. [doi: 10.1002/dmrr.277] [Medline: 12324985]
96. Rebolledo JA, Arellano R. Cultural Differences and Considerations When Initiating Insulin. Diabetes Spectr 2016 Aug 15;29(3):185-190 [FREE Full text] [doi: 10.2337/diaspect.29.3.185] [Medline: 27574375]
97. Tyler NS, Jacobs PG. Artificial Intelligence in Decision Support Systems for Type 1 Diabetes. Sensors (Basel) 2020 Jul 05;20(11):3214 [FREE Full text] [doi: 10.3390/s20113214] [Medline: 32517068]
98. Chaki J, Thillai Ganesh S, Cidham S, Ananda Theertan S. Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review. Journal of King Saud University - Computer and Information Sciences 2020 Jul [FREE Full text] [doi: 10.1016/j.jksuci.2020.06.013]
99. Gamble A, Pham Q, Goyal S, Cafazzo JA. The Challenges of COVID-19 for People Living With Diabetes: Considerations for Digital Health. JMIR Diabetes 2020 May 15;5(2):e19581 [FREE Full text] [doi: 10.2196/19581] [Medline: 32392473]
100. Rubino F, Amiel SA, Zimmet P, Alberti G, Bornstein S, Eckel RH, et al. New-Onset Diabetes in Covid-19. N Engl J Med 2020 Aug 20;383(8):789-790 [FREE Full text] [doi: 10.1056/NEJMc2018688] [Medline: 32530585]
101. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? Lancet Respir Med 2020 Apr;8(4):e21 [FREE Full text] [doi: 10.1016/S2213-2600(20)30116-8] [Medline: 32171062]
102. Guo W, Li M, Dong Y, Zhou H, Zhang Z, Tian C, et al. Diabetes is a risk factor for the progression and prognosis of COVID-19. Diabetes Metab Res Rev 2020 Mar 31:e3319 [FREE Full text] [doi: 10.1002/dmrr.3319] [Medline: 32233013]
103. Chee YJ, Ng SJH, Yeoh E. Diabetic ketoacidosis precipitated by Covid-19 in a patient with newly diagnosed diabetes mellitus. Diabetes Res Clin Pract 2020 Jul;164:108166 [FREE Full text] [doi: 10.1016/j.diabres.2020.108166] [Medline: 32339533]
104. Marchand L, Pecquet M, Luyton C. Type 1 diabetes onset triggered by COVID-19. Acta Diabetol 2020 Oct 11;57(10):1265-1266 [FREE Full text] [doi: 10.1007/s00592-020-01570-0] [Medline: 32653960]
105. Wang S, Ma P, Zhang S, Song S, Wang Z, Ma Y, et al. Fasting blood glucose at admission is an independent predictor for 28-day mortality in patients with COVID-19 without previous diagnosis of diabetes: a multi-centre retrospective study. Diabetologia 2020 Oct 10;63(10):2102-2111 [FREE Full text] [doi: 10.1007/s00125-020-05209-1] [Medline: 32647915]
106. Yang J, Lin S, Ji X, Guo L. Binding of SARS coronavirus to its receptor damages islets and causes acute diabetes. Acta Diabetol 2010 Oct 31;47(3):193-199 [FREE Full text] [doi: 10.1007/s00592-009-0109-4] [Medline: 19333547]
107. Singh AK, Gupta R, Ghosh A, Misra A. Diabetes in COVID-19: Prevalence, pathophysiology, prognosis and practical considerations. Diabetes Metab Syndr 2020 Jul;14(4):303-310 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.004] [Medline: 32298981]
108. CDC. Coronavirus Disease 2019 (COVID-19). Centers for Disease Control and Prevention (2020). URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html [accessed 2021-01-31]

Abbreviations

AI: artificial intelligence
SES: socioeconomic status
T1D: type 1 diabetes
T2D: type 2 diabetes
VBAC: Vaginal Birth after Cesarean

Edited by G Eysenbach; submitted 14.08.20; peer-reviewed by D Gunasekeran, K Gajewska; comments to author 13.10.20; revised version received 02.11.20; accepted 16.01.21; published 10.02.21.

Please cite as:
Pham Q, Gamble A, Hearn J, Cafazzo JA
The Need for Ethnoracial Equity in Artificial Intelligence for Diabetes Management: Review and Recommendations
J Med Internet Res 2021;23(2):e22320
URL: http://www.jmir.org/2021/2/e22320/
doi:10.2196/22320
PMID:33565982

Viewpoint

A Novel Patient Values Tab for the Electronic Health Record: A User-Centered Design Approach

Anjali Varma Desai[1,2], MSCE, MD; Chelsea L Michael[1], MS; Gilad J Kuperman[1], MD, PhD; Gregory Jordan[1], BS; Haley Mittelstaedt[1], MFA; Andrew S Epstein[1,2], MD; MaryAnn Connor[1], MSN, RN-BC, CPHIMS, FAMIA; Rika Paula B Villar[1], BSN, MSHI, RN-BC; Camila Bernal[1], MPH; Dana Kramer[1], NP; Mary Elizabeth Davis[1], RN, DNP, AOCNS; Yuxiao Chen[1], BS; Catherine Malisse[1], BS; Gigi Markose[1], MBA, MSN, RN-BC; Judith E Nelson[1,2], MD, JD

[1]Memorial Sloan Kettering Cancer Center, New York, NY, United States

[2]Department of Medicine, Weill Cornell Medical College, New York, NY, United States

Corresponding Author:
Anjali Varma Desai, MSCE, MD
Memorial Sloan Kettering Cancer Center
1275 York Avenue
New York, NY, 10065
United States
Phone: 1 917 865 2495
Email: Desaia2@mskcc.org

Abstract

Background: The COVID-19 pandemic has shined a harsh light on a critical deficiency in our health care system: our inability to access important information about patients' values, goals, and preferences in the electronic health record (EHR). At Memorial Sloan Kettering Cancer Center (MSK), we have integrated and systematized health-related values discussions led by oncology nurses for newly diagnosed cancer patients as part of routine comprehensive cancer care. Such conversations include not only the patient's wishes for care at the end of life but also more holistic personal values, including sources of strength, concerns, hopes, and their definition of an acceptable quality of life. In addition, health care providers use a structured template to document their discussions of patient goals of care.

Objective: To provide ready access to key information about the patient as a person with individual values, goals, and preferences, we undertook the creation of the Patient Values Tab in our center's EHR to display this information in a single, central location. Here, we describe the interprofessional, interdisciplinary, iterative process and user-centered design methodology that we applied to build this novel functionality as well as our initial implementation experience and plans for evaluation.

Methods: We first convened a working group of experts from multiple departments, including medical oncology, health informatics, information systems, nursing informatics, nursing education, and supportive care, and a user experience designer. We conducted in-depth, semistructured, audiorecorded interviews of over 100 key stakeholders. The working group sought consensus on the tab's main content, homing in on high-priority areas identified by the stakeholders. The core content was mapped to various EHR data sources. We established a set of high-level design principles to guide our process. Our user experience designer then created wireframes of the tab design. The designer conducted usability testing with physicians, nurses, and other health professionals. Data validation testing was conducted.

Results: We have already deployed the Patient Values Tab to a pilot sample of users in the MSK Gastrointestinal Medical Oncology Service, including physicians, advanced practice providers, nurses, and administrative staff. We have early evidence of the positive impact of this EHR innovation. Audit logs show increasing use. Many of the initial user comments have been enthusiastically positive, while others have provided constructive suggestions for additional tab refinements with respect to format and content.

Conclusions: It is our challenge and obligation to enrich the EHR with information about the patient as a person. Realization of this capability is a pressing public health need requiring the collaboration of technological experts with a broad range of clinical leaders, users, patients, and families to achieve solutions that are both principled and practical. Our new Patient Values Tab represents a step forward in this important direction.

(J Med Internet Res 2021;23(2):e21615) doi:10.2196/21615

KEYWORDS

electronic health record; health informatics; supportive care; palliative care; oncology

Introduction

Background

Besides inadequate personal protective equipment and limited intensive care surge capacity, the COVID-19 pandemic has shined a harsh light on another critical deficiency in our system of health care delivery: our inability to access important information about patients' values, goals, and preferences in the electronic health record (EHR) [1-3]. All over the United States, frontline clinicians in the field, emergency departments, intensive care units (ICUs), and on rapid response teams and hospital floors have struggled with urgent decisions about the use of life-supporting technologies to treat serious complications of COVID-19 infection without sufficient information about the patient as a person, what means most to this individual, how the patient defines living well, and whom the patient trusts to make important decisions about health care [4,5]. Such information is rarely accessible even for older adults and others with underlying diseases, such as cancer or chronic comorbid conditions, who are most vulnerable. Either this information was not previously elicited or, although it was discussed, it was not documented or is difficult to find in the EHR [6-9].

This is not a new problem, and it will certainly persist after COVID-19 in the absence of major innovative efforts. Although many clinicians consider the EHR to pose a barrier to patient-centered care by literally shifting the attention from the patient to the computer screen [10-12], the EHR is also a potentially powerful tool that can support clinician-patient communication, team collaboration, personalized and respectful care, continuity across settings, patient engagement, and shared decision making in accordance with patients' individual needs and priorities [13,14]. Several recent initiatives have utilized digital platforms and tools (eg, documentation templates, automated prompts, and electronic order sets) to optimize documentation of advance care planning and goals of care discussions [15-17]. Some have focused primarily on the documentation of advance care planning in patients who are older or have advanced disease [18,19]. A more recent initiative leverages the Epic EHR by assembling information about serious illness conversations, including prognostic information given by clinicians to the patient and family, the patient's understanding of the course of the illness, hopes and worries, priorities, and clinician recommendations, into a new EHR template [20].

At Memorial Sloan Kettering Cancer Center (MSK), we integrated and systematized health-related values discussions led by oncology nurses for newly diagnosed cancer patients as part of routine comprehensive cancer care, regardless of the patient's stage, prognosis, or treatment intent [21,22]. These discussions are revisited quarterly or as deemed appropriate after prespecified clinical events (eg, progression of disease through first-line therapy, hospitalization, or admission to the ICU). In this model, communication encompasses not only the patient's wishes for care at the end of life but also more basic and holistic personal values, including sources of strength, concerns, hopes, definition of an acceptable quality of life, and what the patient wants the clinical team to know about them as a person in order to provide the best care and preserve dignity [22]. In addition, oncologists and other physicians and advanced practice providers use a structured template to document their discussions of patient goals of care, which may address the expected course of the patient's illness, intent of the current treatment, goals identified by the patient, preferences for end-of-life care, and, if relevant, hospice enrollment.

The Patient Values Tab: Concept and Design

To provide ready access to these crucial communications between clinicians and patients as well as key information about the patient as a person with individual values, goals, and preferences, we undertook the creation of the Patient Values Tab in our center's EHR to display this information in a single centralized location. Our institution's EHR platform is Allscripts Sunrise Clinical Manager (Allscripts Healthcare LLC), which organizes data displays into tabs, as do many other EHR platforms. Here, we describe the interprofessional, interdisciplinary, iterative process and user-centered design methodology we applied to build this novel functionality [23].

Methods

Working Group and Stakeholder Interviews

We first convened a working group of experts from multiple departments, including medical oncology, health informatics, information systems, nursing informatics, nursing education, and supportive care (a multidisciplinary palliative care service), and a user experience designer. Members of the working group were selected for their previous experience leading large-scale institutional initiatives and unique expertise in the integration of supportive care and oncology. This core team met every 2 weeks to collaborate on the development and design of the new tab.

To understand the needs and perspectives of a broad range of institutional stakeholders, we also conducted in-depth, semistructured, audiorecorded interviews of over 100 key stakeholders based on a written guide prepared by the core team. Analysis of these interviews, which is reported in a separate publication (which includes selected stakeholder comments in an appendix), was used to inform content and format of the tab [24]. As active contributors to the creative process, those who were interviewed were more inclined to buy in to the ultimate product. In addition, many of these stakeholders held leadership positions within divisions and departments and went on to share enthusiasm about the upcoming tab with their colleagues, enhancing the visibility of the ongoing development effort among a broader group of users.

Mapping Tab Content

The core team sought consensus on the Patient Values Tab's main content, homing in on high-priority areas identified in the stakeholder interviews while also incorporating their input on format and considering suggestions on the logistics of the implementation process. The core content was mapped to various data sources within the EHR (as shown in Multimedia Appendix 1).

For example, oncology nurses use a structured document entitled "Assessment, Patient Personal Values" to summarize their values discussions. The patient's preferred name, language, and communication preferences are elicited through the digital patient portal system via an electronic care questionnaire, which the nurse verifies and updates as needed at the first clinic visit on the nursing health assessment. Education about health care proxy (HCP) and other advance directives is provided in the hospital by patient representatives, who also record designation of and information about the HCP in a specific clinical document.

High-Level Design Principles

As the data sources were clarified, we also established a set of high-level design principles (shown in Textbox 1) to guide this process.

Textbox 1. High-level design principles.

1. The Patient Values Tab should provide users with an at-a-glance understanding of the patient as a person.
2. The Patient Values Tab should offer easy access to data that can be viewed in aggregate in the context of other relevant information to support clinical decision making.
3. The Patient Values Tab will specify the context in which the information was collected (with the source and date the information was updated).
4. The Patient Values Tab will be in a read-only format (ie, the data displayed will be captured and edited elsewhere).
5. The Patient Values Tab will contain the most high-yield information while presenting this information in a succinct, streamlined way to minimize cognitive load for users.
6. The Patient Values Tab will be accessible to all health care team members across the spectrum of patient care.

One such principle was that the tab would be in a read-only display format, populated with existing source documents in the EHR that could not be directly edited in the tab itself. We chose this approach because we and various key stakeholders we interviewed were concerned that the content would become unwieldy and unstable, ultimately leading to inefficiency for users if multiple modifications were allowed. We concluded that reliance on existing workflows and processes would reduce the cognitive load for clinicians and enable a communal responsibility for the underlying documentation, with important roles for various team members as contributors to a shared understanding of the patient as a person.

User Testing

The next step was close collaboration with our user experience designer to create and refine wireframes of the tab design. During this process, the designer conducted usability testing with physicians, nurses, and other health professionals. All participants were asked to provide general usability feedback on the design. To obtain more detailed feedback, participants were presented with specific clinical tasks calling for the use of information contained in the Patient Values Tab in scenarios that the core team generated and adapted for particular roles and responsibilities. For example, in one scenario, the physician was preparing to "discuss serious results with the patient," while the nurse needed to "provide information for medication management." This process revealed a lack of familiarity among users with some of the underlying documentation that sourced the data displayed in the tab. To address this issue, we created a frequently asked questions tile (ie, section) in the tab including basic information about data sources and directing users to the appropriate underlying document to update the information, if needed.

Refinement and Validation

Additional design refinements incorporated (1) suggestions from the user experience designer (eg, creating a banner at the top of the tab with the preferred name information, color coding the different categories of information to help users recognize and navigate among them quickly, including a separate feedback tile within the tab inviting users' input and comments about their experience using the tab); (2) input from the core team (eg, placing the goals of care discussions and the nurse values summary in the most central, prominent positions in the tab; extracting and separately displaying the essential information regarding emergency contact, HCP, and next of kin for rapid access during clinical emergencies, with relevant advance directive scanned forms located below this information); and (3) insights gleaned from the stakeholder interviews (eg, providing expedited access to consultant notes from supportive care, psychiatry, and ethics as the highest yield and most consistently complete sources of information about patient values and personhood). During this iterative process, the designer shared mock-up options and elicited feedback from the core team via email and throughout the series of biweekly core team meetings.

Members of our core team with informatics expertise conducted data validation testing to ensure the fidelity of the information displayed in the tab to the underlying source of this information (ie, confirming that the tab displayed the correct information from the intended source document in the intended format for the correct patient). In addition, to detect delays in displaying information to the user that might decrease usability, we assessed the time to launch the tab content for selected patients. If time delays were identified, we worked with the information systems team to identify the source of the delay so that it could

be rectified; specifically, technical optimizations were made to bring the load times in line with other feature load times in our EHR (ie, no longer than a few seconds).

Results

Initial Deployment and Implementation

We have already deployed the Patient Values Tab to a pilot sample of users in the Gastrointestinal (GI) Medical Oncology Service at our center, including the physicians, advanced practice providers (APPs), nurses, and administrative staff. We chose this service for this first phase of our pilot deployment because it is the largest solid tumor service at MSK, with approximately 40 physicians varying in age, gender, race and ethnicity, number of years in clinical practice, and disease focus who care for a large number of cancer patients with diverse demographics (eg, age, gender) and types and stages of cancer with wide variability in the pace of the disease process and overall clinical course. Patients of the GI Medical Oncology Service represent the largest proportion of admissions to our Memorial Hospital, thus frequently involving both outpatient and inpatient teams in their care.

We have executed a detailed implementation plan, which includes (1) discussion with the physician, APP, and nursing leaders of the service about the tab and its integration in clinical care by that service; (2) email notification to users about the availability and basic content of the new tab; and (3) presentations encompassing key features and functionality of the tab together with options for incorporating its use in regular clinical workflows. For each of these steps, we have targeted the professional groups individually, since they have different needs, roles, and workflows in providing patient care, although ultimately, they must all collaborate and communicate as a team to optimize this care.

Further Iteration and Evaluation

Learnings from this pilot will inform further refinement of the Patient Values Tab before broader implementation. We are measuring usage of the tab through audit logs, which specify the time, user, and patient involved each time it is launched. In addition, we have included a feedback tile within the tab itself that asks in closed-ended items for users' broad impressions ("It's great," "It's okay," or "It needs improvement") and provides space for free-text comments. In-depth feedback will be gathered from selected clinicians through brief individual interviews. We plan to select these clinicians based on audit log data indicating those who are low adopters or high adopters in terms of the frequency and timing of their tab use. Through the interviews, we will explore whether the tab fits with the clinician's workflow and patterns of use (or lack of use) in clinical practice, the perceived value of the tab to the clinician, and suggestions for improvement (eg, with respect to content and format).

Guided by this diverse user input, we have continued cycles of refinement and deployment to additional groups incrementally, with a plan for institution-wide rollout by the fall of 2020. The core team is continuing its biweekly meetings, monitoring every step in the implementation and evaluation process, and collaborating to address and incorporate user feedback in order to ensure that the use of the tab is maximized and sustained.

Early Findings

We already have early evidence of the positive impact of this EHR innovation. Audit logs show increasing use among different user groups. Among users who have engaged with the feedback section within the tab, 25 of 48 (52%) users reported that "it is great" compared with 18 of 48 (38%) users who reported that "it needs improvement" and provided specific constructive suggestions and 5 of 48 (10%) users who reported that "it is okay." Many of these initial user comments have been enthusiastically positive (Textbox 2).

Textbox 2. Illustrative quotes.

"[I] love the easy access to [health care proxy] and Advance directives!"

"Being able to quickly check the patient's preferred name is incredibly helpful!"

"I like that it captures information from many different areas and placed in one area. I will be using this tab."

"Numerous tiles you have developed give the patient a 'voice' in our [electronic health record]. Strong work….This will be a great advance in Cognitive Support for our clinical teams."

Users also suggested specific enhancements to the tab's content quality and design.

Our early pilot data have also revealed potential barriers to large-scale implementation, which include perceived lack of time to engage with the tab, competing priorities, incomplete or inaccurate tab content, technical glitches, and difficulty remembering to launch the tab as part of routine workflow. To address these challenges, we have enacted several proactive educational strategies when deploying to additional user groups, including (1) emphasizing that the tab is a display-only feature that is intended to save time by consolidating key information in a central location, (2) describing the processes by which incomplete or inaccurate tab content can be updated by the user, and (3) highlighting the "share feedback" section as a means for providing feedback on technical (or content-related) glitches that can be addressed by the Patient Values Tab interdisciplinary working group. In all circumstances, the Patient Values Tab working group follows up with each individual user who provides feedback to ensure closed-loop communication and enhance transparency in the ongoing effort to refine and optimize the tab's functionality.

We are also increasingly receiving reports of specific patient situations in which the information displayed in the tab, including the outpatient oncology nurse's values summary, allowed the hospital team to support families in difficult decisions about the use of intensive care therapies. In one such

case, the family described hearing the ICU physician reading the values summary aloud as feeling that the patient, who no longer was conscious and had not discussed this type of situation with them previously, "was in the room, speaking directly to them," clarifying his priorities, providing guidance, and relieving them of burden and guilt in deciding to limit life support at that time. Immediate access that the Patient Values Tab provides to primary physicians' goals of care discussions greatly facilitates the work of our emergency department and rapid response team clinicians, who must act quickly in emergencies.

Similarly, we are hearing that the communication preferences section within the Patient Values Tab is particularly helpful to teams as they are preparing for family meetings and goals of care discussions. For example, one Patient Values Tab user recounted:

> *I did use it actually this week.…We were going to go have an impromptu family meeting and the attending physician asked, "Would the patient want to be involved?" and I said, "Well, let's look at the Patient Values Tab"….We clicked there and it said "patient wants a lot of information with her partner present"…so…we used it in real time.*

In this case, knowing how the patient preferred to receive medical information (ie, in detail) and with whom present (ie, her partner) helped guide the health care team's approach to having a goals of care conversation at a timely moment.

Discussion

Even at this pilot stage of deployment, it is becoming clear that the Patient Values Tab enables care and decision making that honors the personhood and values of our patients. During the COVID-19 pandemic, as life-threatening emergencies have occurred without warning to patients who were then too ill to speak for themselves, we have learned—again, in a more painful but perhaps more enduring lesson—that high-quality care is heavily dependent on immediate access to this crucial information. As implementation and usage of the tab expands and our center moves beyond the current COVID-19 crisis, we will be able to examine its impact on a variety of outcomes at the level of the patient, clinician, work process, and health care system. Although the Patient Values Tab was built in MSK's specific EHR (ie, Allscripts), this software can be configured in other EHR platforms as well, including Epic (the predominant platform in the United States). Further research is needed on the best ways to optimize and enhance the patient centeredness of various EHR systems through the synergistic integration of tools that capture broader, holistic patient values (eg, the Patient Values Tab), with ongoing efforts that are primarily focused on advance care planning (eg, Epic's Advance Care Planning tab).

All EHRs contain extensive quantitative information and voluminous data about patients that are mostly impersonal. Now, it is our challenge and obligation to enrich the EHR with information about the patient as a person, which is rarely included or readily accessible. Patient-centered care is highly prioritized by the Institute of Medicine [25], patients and families, and professional caregivers [26], but it can only be delivered if the patient's personal values are as prominent as the laboratory values in the EHR on which health care professionals rely and spend the bulk of their time. Although it can be burdensome and distracting for clinical care, the EHR has tremendous untapped potential to support patient-centered care. Realization of this capability is a pressing public health need requiring the full collaboration of technological experts with a broad range of clinical leaders, users, patients, and families to achieve solutions that are both principled and practical. Our new Patient Values Tab is a step forward in this important direction.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table S1. Mapping the Patient Values Tab content to source documentation.
[DOCX File , 20 KB - jmir_v23i2e21615_app1.docx]

References

1. Sanders JJ, Paladino J, Reaves E, Luetke-Stahlman H, Anhang Price R, Lorenz K, et al. Quality Measurement of Serious Illness Communication: Recommendations for Health Systems Based on Findings from a Symposium of National Experts. J Palliat Med 2020 Jan;23(1):13-21 [FREE Full text] [doi: 10.1089/jpm.2019.0335] [Medline: 31721629]
2. Prater L, Sanchez A, Modan G, Burgess J, Frier K, Richards N, et al. Electronic Health Record Documentation Patterns of Recorded Primary Care Visits Focused on Complex Communication: A Qualitative Study. Appl Clin Inform 2019 Mar;10(2):247-253 [FREE Full text] [doi: 10.1055/s-0039-1683986] [Medline: 30970382]
3. Varpio L, Rashotte J, Day K, King J, Kuziemsky C, Parush A. The EHR and building the patient's story: A qualitative investigation of how EHR use obstructs a vital clinical activity. Int J Med Inform 2015 Dec;84(12):1019-1028 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.09.004] [Medline: 26432683]
4. Curtis JR, Kross EK, Stapleton RD. The Importance of Addressing Advance Care Planning and Decisions About Do-Not-Resuscitate Orders During Novel Coronavirus 2019 (COVID-19). JAMA 2020 May 12;323(18):1771-1772. [doi: 10.1001/jama.2020.4894] [Medline: 32219360]

5. Lee J, Abrukin L, Flores S, Gavin N, Romney M, Blinderman CD, et al. Early Intervention of Palliative Care in the Emergency Department During the COVID-19 Pandemic. JAMA Intern Med 2020 Sep 01;180(9):1252-1254 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2713] [Medline: 32501486]
6. Lamas D, Panariello N, Henrich N, Hammes B, Hanson LC, Meier DE, et al. Advance Care Planning Documentation in Electronic Health Records: Current Challenges and Recommendations for Change. J Palliat Med 2018 Apr;21(4):522-528. [doi: 10.1089/jpm.2017.0451] [Medline: 29360417]
7. Curtis JR, Sathitratanacheewin S, Starks H, Lee RY, Kross EK, Downey L, et al. Using Electronic Health Records for Quality Measurement and Accountability in Care of the Seriously Ill: Opportunities and Challenges. J Palliat Med 2018 Mar;21(S2):S52-S60. [doi: 10.1089/jpm.2017.0542] [Medline: 29182487]
8. Walker E, McMahan R, Barnes D, Katen M, Lamas D, Sudore R. Advance Care Planning Documentation Practices and Accessibility in the Electronic Health Record: Implications for Patient Safety. J Pain Symptom Manage 2018 Feb;55(2):256-264 [FREE Full text] [doi: 10.1016/j.jpainsymman.2017.09.018] [Medline: 28943360]
9. Lakin JR, Isaacs E, Sullivan E, Harris HA, McMahan RD, Sudore RL. Emergency Physicians' Experience with Advance Care Planning Documentation in the Electronic Medical Record: Useful, Needed, and Elusive. J Palliat Med 2016 Jun;19(6):632-638. [doi: 10.1089/jpm.2015.0486] [Medline: 27203483]
10. Makoul G, Curry RH, Tang PC. The use of electronic medical records: communication patterns in outpatient encounters. J Am Med Inform Assoc 2001;8(6):610-615 [FREE Full text] [doi: 10.1136/jamia.2001.0080610] [Medline: 11687567]
11. Rathert C, Mittler JN, Banerjee S, McDaniel J. Patient-centered communication in the era of electronic health records: What does the evidence say? Patient Educ Couns 2017 Jan;100(1):50-64. [doi: 10.1016/j.pec.2016.07.031] [Medline: 27477917]
12. Street RL, Liu L, Farber NJ, Chen Y, Calvitti A, Zuest D, et al. Provider interaction with the electronic health record: the effects on patient-centered communication in medical encounters. Patient Educ Couns 2014 Sep;96(3):315-319 [FREE Full text] [doi: 10.1016/j.pec.2014.05.004] [Medline: 24882086]
13. White A, Danis M. Enhancing patient-centered communication and collaboration by using the electronic health record in the examination room. JAMA 2013 Jun 12;309(22):2327-2328 [FREE Full text] [doi: 10.1001/jama.2013.6030] [Medline: 23757080]
14. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 02;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
15. Karim S, Harle I, O'Donnell J, Li S, Booth CM. Documenting Goals of Care Among Patients With Advanced Cancer: Results of a Quality Improvement Initiative. J Oncol Pract 2018 Sep;14(9):e557-e565. [doi: 10.1200/JOP.18.00031] [Medline: 30113873]
16. Huber MT, Highland JD, Krishnamoorthi VR, Tang JW. Utilizing the Electronic Health Record to Improve Advance Care Planning: A Systematic Review. Am J Hosp Palliat Care 2018 Mar;35(3):532-541. [doi: 10.1177/1049909117715217] [Medline: 28627287]
17. Haley EM, Meisel D, Gitelman Y, Dingfield L, Casarett DJ, O'Connor NR. Electronic Goals of Care Alerts: An Innovative Strategy to Promote Primary Palliative Care. J Pain Symptom Manage 2017 May;53(5):932-937 [FREE Full text] [doi: 10.1016/j.jpainsymman.2016.12.329] [Medline: 28062333]
18. Turley M, Wang S, Meng D, Kanter M, Garrido T. Impact of a Care Directives Activity Tab in the Electronic Health Record on Documentation of Advance Care Planning. Perm J 2016;20(2):43-48 [FREE Full text] [doi: 10.7812/TPP/15-103] [Medline: 27057820]
19. Paladino J, Bernacki R, Neville BA, Kavanagh J, Miranda SP, Palmor M, et al. Evaluating an Intervention to Improve Communication Between Oncology Clinicians and Patients With Life-Limiting Cancer: A Cluster Randomized Clinical Trial of the Serious Illness Care Program. JAMA Oncol 2019 Jun 01;5(6):801-809. [doi: 10.1001/jamaoncol.2019.0292] [Medline: 30870556]
20. Wilson E, Bernacki R, Lakin JR, Alexander C, Jackson V, Jacobsen J. Rapid Adoption of a Serious Illness Conversation Electronic Medical Record Template: Lessons Learned and Future Directions. J Palliat Med 2020 Feb;23(2):159-161. [doi: 10.1089/jpm.2019.0420] [Medline: 32023189]
21. Desai AV, Klimek VM, Chow K, Epstein AS, Bernal C, Anderson K, et al. 1-2-3 Project: A Quality Improvement Initiative to Normalize and Systematize Palliative Care for All Patients With Cancer in the Outpatient Clinic Setting. J Oncol Pract 2018 Dec;14(12):e775-e785. [doi: 10.1200/JOP.18.00346] [Medline: 30537456]
22. Epstein AS, Desai AV, Bernal C, Romano D, Wan PJ, Okpako M, et al. Giving Voice to Patient Values Throughout Cancer: A Novel Nurse-Led Intervention. J Pain Symptom Manage 2019 Jul;58(1):72-79.e2 [FREE Full text] [doi: 10.1016/j.jpainsymman.2019.04.028] [Medline: 31034869]
23. Abras C, Maloney-Krichmar D, Preece J. User-Centered Design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Thousand Oaks, CA: Sage Publications; 2004.
24. Desai AV, Agarwal R, Epstein AS, Kuperman GJ, Michael CL, Mittelstaedt H, et al. Needs and Perspectives of Cancer Center Stakeholders for Access to Patient Values in the Electronic Health Record [published online ahead of print, 2021 Feb 8]. JCO Oncol Pract 2021:OP2000644. [doi: 10.1200/OP.20.00644] [Medline: 33555928]

25. Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
26. Gerteis M, Edgman-Levitan S, Daley J, Delbanco TL, editors. Through the Patient's Eyes: Understanding and Promoting Patient-Centered Care. San Francisco, CA: Jossey-Bass; 1993.

Abbreviations

APP: advanced practice providers
EHR: electronic health record
GI: gastrointestinal
HCP: health care proxy
ICU: intensive care unit
MSK: Memorial Sloan Kettering Cancer Center

Edited by G Eysenbach, R Kukafka; submitted 14.07.20; peer-reviewed by S Bose-Brill, R Lee; comments to author 01.08.20; revised version received 21.08.20; accepted 21.10.20; published 17.02.21.

Please cite as:
Desai AV, Michael CL, Kuperman GJ, Jordan G, Mittelstaedt H, Epstein AS, Connor M, B Villar RP, Bernal C, Kramer D, Davis ME, Chen Y, Malisse C, Markose G, Nelson JE
A Novel Patient Values Tab for the Electronic Health Record: A User-Centered Design Approach
J Med Internet Res 2021;23(2):e21615
URL: http://www.jmir.org/2021/2/e21615/
doi:10.2196/21615
PMID:33595448

Viewpoint

Defining Digital Game-Based Learning for Science, Technology, Engineering, and Mathematics: A New Perspective on Design and Developmental Research

Shahrul Affendi Ishak[1], BAA, MA; Rosseni Din[1], AAD, BSc, MEd, PhD; Umi Azmah Hasran[2], BSc, MSc, PhD

[1]Science, Technology, Engineering, and Mathematics Enculturation Research Centre, Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia

[2]Fuel Cell Institute, Universiti Kebangsaan Malaysia, Bangi, Malaysia

Corresponding Author:
Rosseni Din, AAD, BSc, MEd, PhD
Science, Technology, Engineering, and Mathematics Enculturation Research Centre
Faculty of Education
Universiti Kebangsaan Malaysia
43600 UKM
Bangi, 43600
Malaysia
Phone: 60 166656420
Email: rosseni@ukm.edu.my

Abstract

In the modern age, digital games are widely used as informal media for Science, Technology, Engineering, and Mathematics (STEM) education and medical therapy for game-based learning. Digital games provide learners with a graphical system of interaction that enhances scientific concepts within an enjoyable environment. The vastly increasing number of digital games produced in the market affects the quality of STEM digital games while requiring multidisciplinary expertise. This paper proposes a framework for STEM digital game-based learning encompassing input-process-output stages. Several studies from the early 2000s onward were reviewed to discuss and present a new perspective on a framework for the design and development of digital games, particularly for STEM. This proposed framework consists of digital game development as input, experience as a process, and constructs as output. This simple and precise framework will generate a universal product for various types of learners. It can thus be used as a guideline for game designers, developers, and experts to develop STEM digital games and achieve better learning outcomes.

(J Med Internet Res 2021;23(2):e20537) doi:10.2196/20537

KEYWORDS

digital game-based learning; STEM digital game; game development model; game design; design and developmental research

Introduction

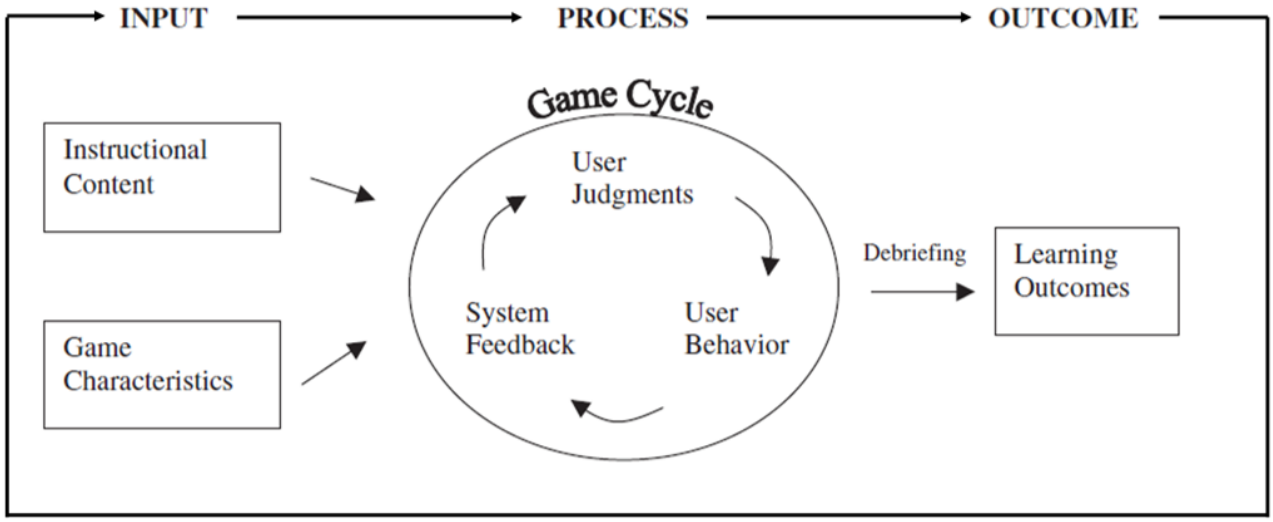
The use of digital games for education has been identified as one of the global pedagogical approaches required for 21st century learners [1]. The intention is to make STEM (science, technology, engineering, and mathematics) education more interactive and interesting and enhance understanding of STEM concepts [2,3]. Individuals in any age range can learn through digital games. Ultimately, it is more practical to use such games in kindergarten to grade 12 education (primary and secondary school) to increase students' levels of interest in STEM. An interactive gaming system uses advanced graphics and programming tools. STEM content can be gamified easily in the current digital era. Game designers and developers often collaborate with teachers and experts to construct good instructional games for STEM learning as a universal product. This type of pedagogical approach is known as digital game-based learning [4] or STEM digital game-based learning [5,6].

Using digital games to enhance STEM education is important. Their purpose is to allow scientific concepts to be visualized [7-9]. The STEM curriculum consists of multiple facts and concepts learners should understand and practice in their daily lives. Some may not be easy to learn using conventional teaching pedagogy [10]. Teachers need supporting learning material to visualize each concept, for which digital games can be the best tools. Digital games allow learners to interact with game mechanics in a virtual world, achieving the set goals as a result

of the desire to win triggered by these games. This makes learners feel enjoyment while simultaneously enabling them to gain a better understanding of STEM content. Their interaction and involvement in experiencing and understanding the game while resolving problems in the virtual world provide a meaningful gaming experience [2,8,9]. Thus, the use of digital games not only enhances learning, develops skills, and increases the memorization and understanding of STEM, it also helps to maintain interest and create a STEM mindset [11].

Digital games for formal and informal education are accessible through computers, tablets, and smartphones, all of which are popular among digital natives—the new digital generation [12]. STEM education is challenging for young people. Most digital games aim to enhance STEM learning on various gaming platforms [13]. This was introduced in the earliest model of game-based learning developed by Garris et al [13], which consisted of input, process, and output (Figure 1). Current studies on gaming and learning have 3 major and distinct perspectives—(1) research on input: how to develop educational digital games; (2) research on process: how the games work for users; and (3) research on outcome: impact of the games. Advancing gaming technology means the number of digital games keeps increasing, and developers are competing with each other [14]. Developing digital games for STEM involves a multitude of team experts. These collaborations lead to an agile form of development that requires multiple lightweight processes over a shorter period. This makes it easier for each game designer, game developer, and STEM expert to ensure the digital game undergoes a systematic development process suitable for STEM digital game-based learning.

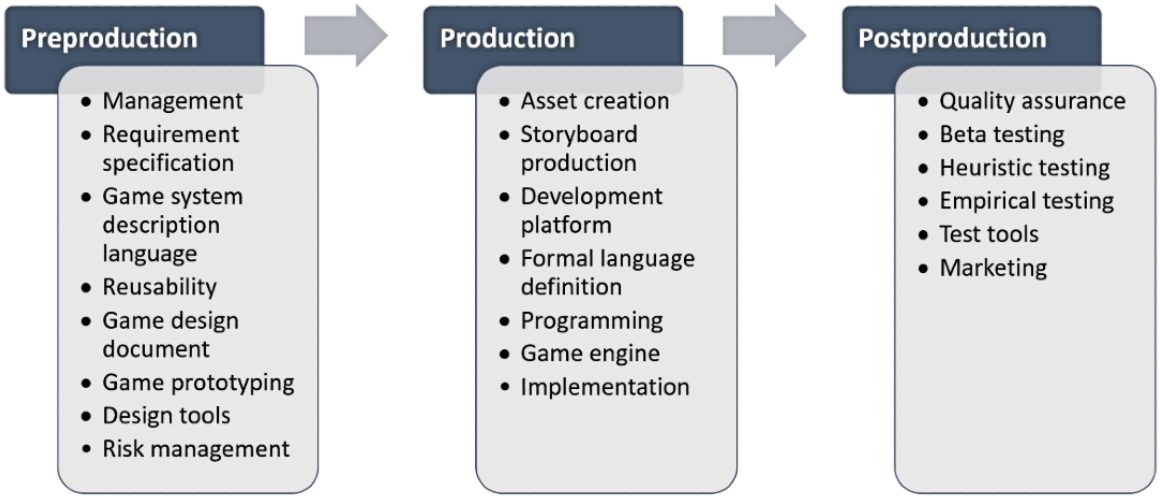
Figure 1. Input-process-output model adapted by Garris et al [13].



Game development processes are integrated and mapped into the design and development research, with emphasis on a systematic process of product development to achieve STEM learning outcomes. Digital game development undergoes iteration processes. From an industrial perspective, this involves preproduction, production, and postproduction phases as a linear process (Figure 2) [15]. However, studies [16-25] involving design and development research indicate that research products should undergo iteration processes until the development objective is achieved. Thus, digital games must be created for use by numerous people regardless of their abilities and demographic characteristics. This concept is synonymous with personalized learning based on a universal design for learning [16,17]. From an educational perspective, digital games are viewed as a learning tool in which STEM content is important in ensuring each game enhances learning and interest in STEM. Although previous studies [2,4,5,8,10,12,13,26-32] have introduced various models of digital game-based learning for specific outcomes, they have not described the development input-process-output structure as a holistic framework.

In the fourth industrial revolution environment, daily human activities have been influenced by digital technology. The new digital generation play digital games as part of their leisure activities. Numerous studies [26,27] on digital games have been conducted over the past two decades and have had a major influence on individual learning process; however, they have rarely discussed STEM learning. Because the number of digital games for STEM keeps increasing as a result of rapid development by the gaming industries, a relevant theoretical and practical underpinning of the development process for better STEM learning outcomes is required. As a universal product, the use of digital games for STEM does not rely on achieving better learning outcomes alone, it also needs to stimulate young people's interest in STEM as a future career from an early age. Thus, this paper proposes a framework to understand STEM digital game-based learning encompassing design and development, gaming experience, and the generation of STEM outcomes.

Figure 2. Game development process from industrial practice.



Digital Game-Based Learning for STEM

Discussions on the use of digital games for STEM can be confusing. It is a multidisciplinary field consisting of 2 different domains: STEM education and digital gaming. The link between these fields triggers the need for further research to understand how and why digital games affect STEM learning. Thus, the use of digital games and STEM learning should be studied in parallel. Digital game-based learning for STEM is also a difficult concept to define. In terms of their separate fields, digital game-based learning is defined as the use of digital games to enhance learning. Thus, to develop a more holistic perspective, we define STEM digital game-based learning as the process of creating an interactive STEM environment through digital games to enhance individual STEM learning. As such, several relevant design and development studies are now reviewed.

Most of the development models for any product undergo a systematic process. Several studies proposed a new model adapted from the original ADDIE (Analysis, Design, Development, Implementation, and Evaluation) model [20,33,34]. Din [18] revealed that the ADDIE model did not consist of a usability test, which is the most important process in product development. Other scholars have also adapted and improved the model to ensure a better systematic process [20,33,34]; however, in the context of developing a digital game, every process undergoes rapid and, some may argue, agile development. This is due to the higher number of digital games produced each day. Gaming industries tend to develop digital games in a shorter time, and some might not follow the systematic process proposed by the literature. This increases the number of low-quality digital games for users.

Din initially developed the Model *Pembangunan Sistem* to design and create a computer conferencing system [18] and further developed the model into Model *RekaBangun Sistem Pengajaran dan Pembelajaran* [19] after testing it in the development and validation of integrated meaningful hybrid e-training for computer science—a theoretical- and empirical-based design and development approach [20]. This model was tested in numerous other studies to generate the fourth version, the Model *RekaBangun Sistem Pengajaran dan Pembelajaran IV* now called the Model *RekaBangun SPP IV* [19]. Between 2011 and 2015, Din et al [16] rigorously tested the model and developed the Universal Design for Learning Instructional Design model. Between 2015 and 2020, Din combined this model with a simpler agile version of the Model *RekaBangun SPP IV* after rigorous testing on more than 15 online and mobile modules, systems, and apps [17]. This final and latest version comes complete with a universal design for learning and agile development method and contains modeling stages with value integration [17,21].

Input as Digital Game Development

In digital game design, designers first identify the input elements. Universal design ensures a high degree of usability and accessible digital games for various types of learners regardless of ability [28,35,36]. Several models have been analyzed and proposed to determine which specific attributes are needed. A game perspective emphasizes content and game characteristics for enjoyment, while educational technology emphasizes learning content and a pedagogical approach. The combination of these 2 perspectives produces a better attribute as input for the game design system.

Gamification is well-known in educational technology as defining the process of integrating learning content into game mechanics [37]. The purpose is to make the learning content more interesting and engaging and to motivate learners throughout all elements of game-design [38]. Studies [6,13,17,19,22,28-30,39,40] indicate that several universal design attributes need to be incorporated into educational games such as learning theory, learning strategy, pedagogy, learning content, and game characteristics such as value integration. For STEM digital games, *STEM learning content* is a more suitable term representing the learning content attribute. All these attributes are subsequently included in the game design system.

Input for digital game development requires several attributes to undergo systematic development processes. These are integrated from digital games and education perspectives [17,19,30,39-43]. Din's early development model [18-20] did not separate instructional design from the development model. Instead, instructional design processes were embedded into

phases of development [19]. After 2015, Din extracted instructional design components from the development phases. This yielded 2 separate models, the instructional design model (Eclectic Universal Design for Learning model) presented in Figure 3 and the development model presented in Figure 4 [17,44]. The learning design is inspired by the universal design for learning model. It represents the expansion of phase 1 through phase 3 of the development model. The original model consisted of 4 main components: (1) eclectic learning theory, (2) eclectic content, (3) eclectic pedagogy, and (4) eclectic learning strategy.

Figure 3. Eclectic Universal Design for Learning model.

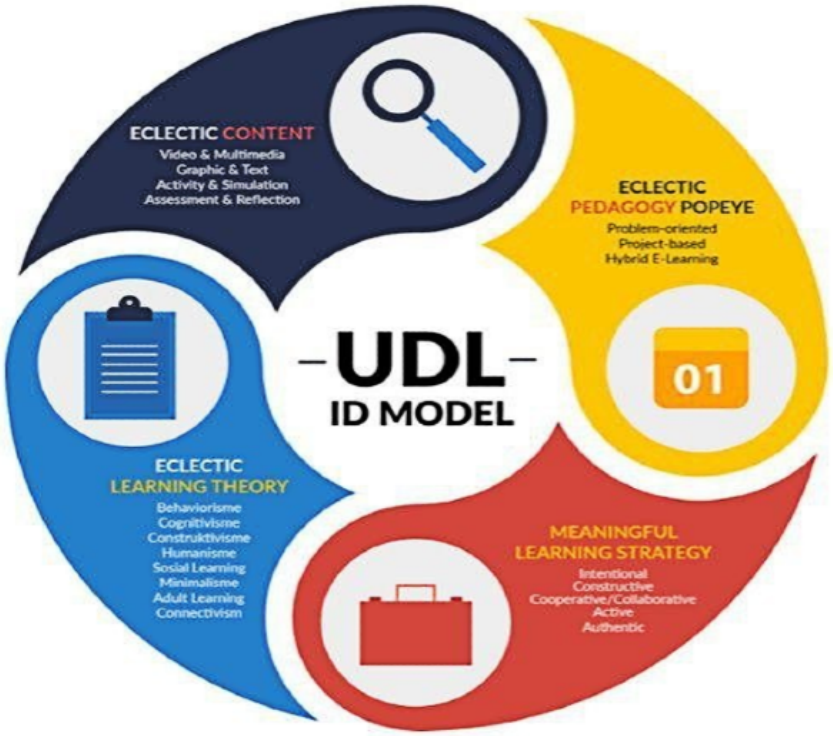
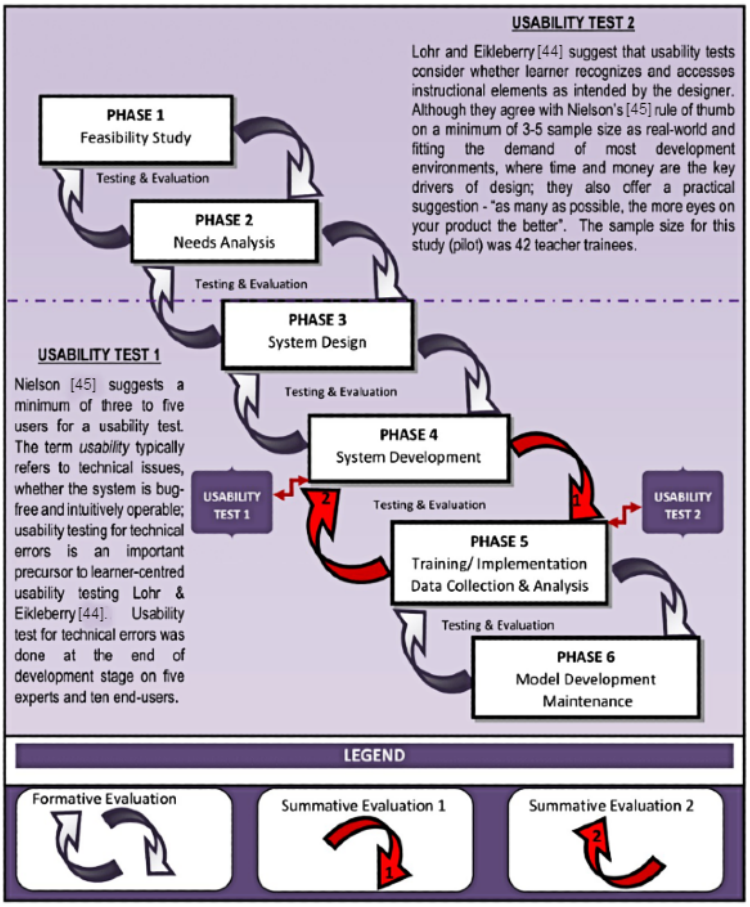
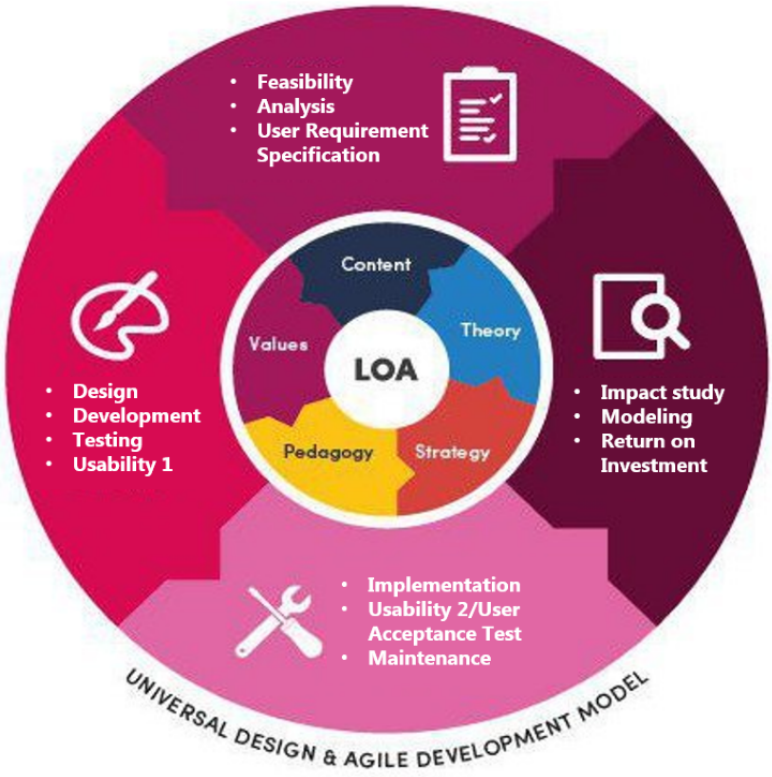


Figure 4. The development model [44,45].



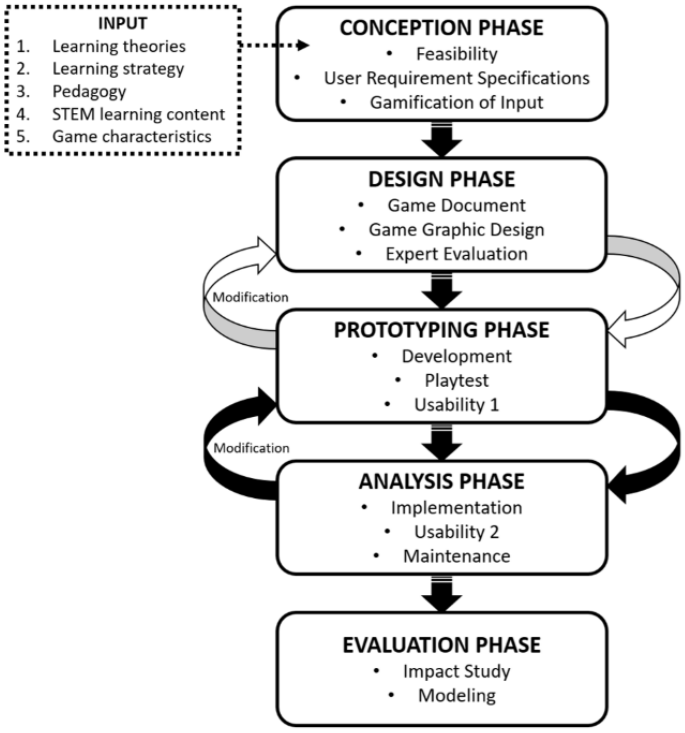
The agile version [17,46] of the integrated model for the instructional design, development, and modeling of a personalized learning environment in education, namely the UDin model, is presented in Figure 5 [46]. According to Din, the UDin model, which is a 20-year transformation model of the design, development, and modeling of a learning system, added the taken-for-granted Learning Outcome component. This was aligned with the Assessment component and both were placed in the center as the innermost part of the model labeled Learning Outcome and Assessment [17,46]. The model emphasizes continuous assessment. The rubrics are mainly used as assessment tools. The assessment methods range from gamification, reflection, and visual-, video- and technology-mediated communication to field work [17,46]. Din's previous work discusses the assessment method and tools in detail; this encompasses studies by Din et al [23-25,47], Azizul and Din [48], Batainah et al [49], Salleh et al [50], and Abdul Manaf et al [51]. Conventional quizzes transformed into interactive online quizzes are also used as formative evaluation alternatives in online modules. Some courses also retain the pen-and-pencil test, primarily for final summative evaluation.

Figure 5. The final UDin model—2 decade transformation of the design, development, and modeling of a learning system.



The simplified version of the STEM digital game development process is presented in Figure 6. As a product for research purposes, the framework consists of 5 main phases: conception phase, design phase, prototyping phase, analyzing phase, and evaluation phase. The conception phase starts with the feasibility of the proposed idea or game concept. Research needs to be undertaken by the developer or game designer to ensure the novelty of the idea compared with existing games. The game developer and STEM experts must therefore examine, analyze, and discuss the main concept of the game that needs to be developed. All 5 components of the universal design for learning mentioned in the input session adapting the UDin model [17,46] are followed in this conception phase. The adapted universal input attributes are learning theories, learning strategy, pedagogy, STEM learning content, and game characteristics. These are gamified into the mechanics of the game. Analysis of user requirement specifications ensures the game has a targeted audience. All discussion regarding the concept is contained in a game specification document that consists of team members, targeted users, game theme, game platform, game mechanics, and game concept.

Figure 6. Proposed STEM digital game development process.



The integration of input through gamification relies on theory. In the proposed model (Figure 6), learning theories serve to frame the context of STEM digital games. Several theories have been associated with digital game-based learning. However, to purposefully create STEM digital games for STEM education, 3 major learning theories have been identified: the theory of experiential learning, theory of self-determination, and the educational-psychology theory of interest development. Blending these theories with a learning strategy, pedagogy, STEM learning content, and game characteristics generates a holistic universal attribute to produce one digital game underlying learning and game development perspectives. The important element in creating STEM digital games is the integration of STEM content. This attribute plays a major role in characterizing STEM digital games. Next, the development of game graphic design is designed based on game specification document and later must to be evaluated by experts.

The prototyping phase is one in which the design and development of the digital games begins. The identified input attributes are ready for the graphics and programming process. All attributes act as an underlying basic or game design in a form of system and mechanics. To program the game system, game designers and programmers follow user-need specifications for the game design. The first prototype of the designed digital games undergoes a playtest among team members prior to a small-scale usability test with a targeted audience. This is to ensure that the digital game works without any major errors. Comments at this point are recorded for improvement. The prototyping phase continues with Usability Test 1 comprising a targeted audience of 3 to 5 users [45]. The games are then modified and improved based on the results.

The next phase is the analysis phase, in which the game is ready for large-scale testing. This consists of Usability Test 2, and all comments are recorded. However, if the result shows another emerging error, the game needs to undergo another prototyping phase so that improvements can be made. The development team need to ensure the iteration process between the prototyping and analysis phases meets the dateline and cost requirements. After the game is tested and data are collected, the programmer makes final improvements during the maintenance phase. The final phase is the evaluation phase in which the performance of the digital game is assessed. The evaluation measures the relevant constructs involved. It also determines whether the game has achieved the objective of development. Due to budget problems, the evaluation process sometimes starts simultaneously with Usability Test 2 [19,22]. In such cases, Usability Test 2 is conducted with a larger targeted audience so that evaluation can be performed at the end of the test.

Process as Experience

User experience is one of the main components of using digital games in STEM education. During testing, there are interactions among users. Users enter the game cycle and make judgements based on observations to understand the system during the first attempt. The behavior of the game provokes the user to try what they have observed regarding the interactions taking place in the game system. The trial-and-error process allows formulation of a meta-framework of conceptual understanding as to how the system works in the developed digital game. The system feedback offered by the digital game due to the designed rules enables users to understand when certain actions do not produce

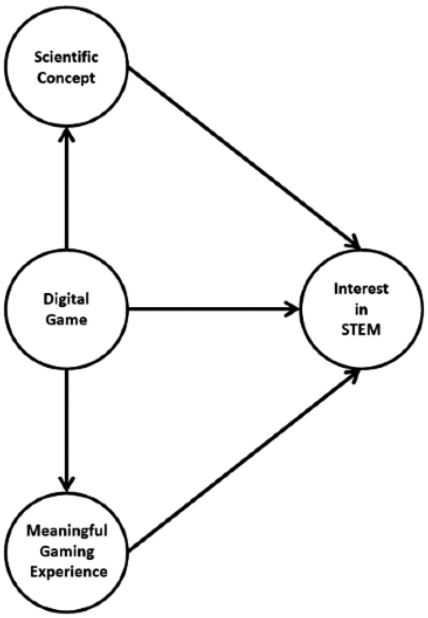
the correct outcome. The cycle occurs continuously while users remain in the gaming world system.

The game cycle and STEM education are related. STEM education emphasizes the need for scientific thinking [11,52,53]. The digital game thus stimulates scientific thinking via observation (understanding how the game system works at the first attempt as a form of assumption), testing (making predictions and testing the assumption as a form of action), and drawing conclusions (system feedback shows whether the action performed produces the current movement inside the game). This game cycle provides the experience of engagement and involvement in an enjoyable environment. At the same time, the digital game helps users practice scientific skills without even realizing they are doing so.

Output as Constructs Involved

There are several constructs involved when the user plays and interacts with the developed digital game. The constructs relate to the main objective of the digital game, which is designed according to STEM purposes. Research shows that any digital game provokes learning due to interaction via the game cycle, investigating the relationship between those constructs. Based on a review of literature, this paper identifies 4 major constructs:(1) digital game, (2) scientific concept, (3) meaningful gaming experience, and (4) interest in STEM (Figure 7). Figure 7 also indicates that the digital game influences interest in STEM through the presence of scientific concepts and meaningful gaming experience as mediators. The major constructs proposed in this paper need to be measured to produce the structural model and measure the effectiveness of STEM digital games.

Figure 7. Proposed relationship of the constructs involve by using digital game.



To determine this relationship, several studies [2,9,11,31,32,54-79] were reviewed and the result mapped into one framework, as depicted in Figure 7. Because the data and results presented are scattered, this model purposely maps the result into the context of STEM digital games (Table 1). Based on the common findings, the digital game has a strong relationship with learning outcomes as the formation of scientific concepts is based on the game content. At the same time, research on digital games as entertainment also reveals a relationship with the gaming experience as the player feels a strong connection with the character inside the game world. However, Fisher [11] and other information sources [1,2,8,9,12,26,27,55-58,79] from the internet claim that the use of a digital game is the best pedagogical approach to increase or stimulate STEM interest among individuals. Digital games may therefore affect individual psychology in that the experience of playing is related to an interest in STEM. This relationship has been confirmed by Krapp [72] where individual interest is a process that undergoes 4 stages [72]. Providing STEM digital games is one of the practices based on Krapp's concept of interest development. According to the literature, stimulating STEM interest using digital games is not a direct cause-effect relationship as another 2 variables have been identified as mediating this effect.

Table 1. Summary of research findings on the relationship between potential constructs as output for STEM digital games.

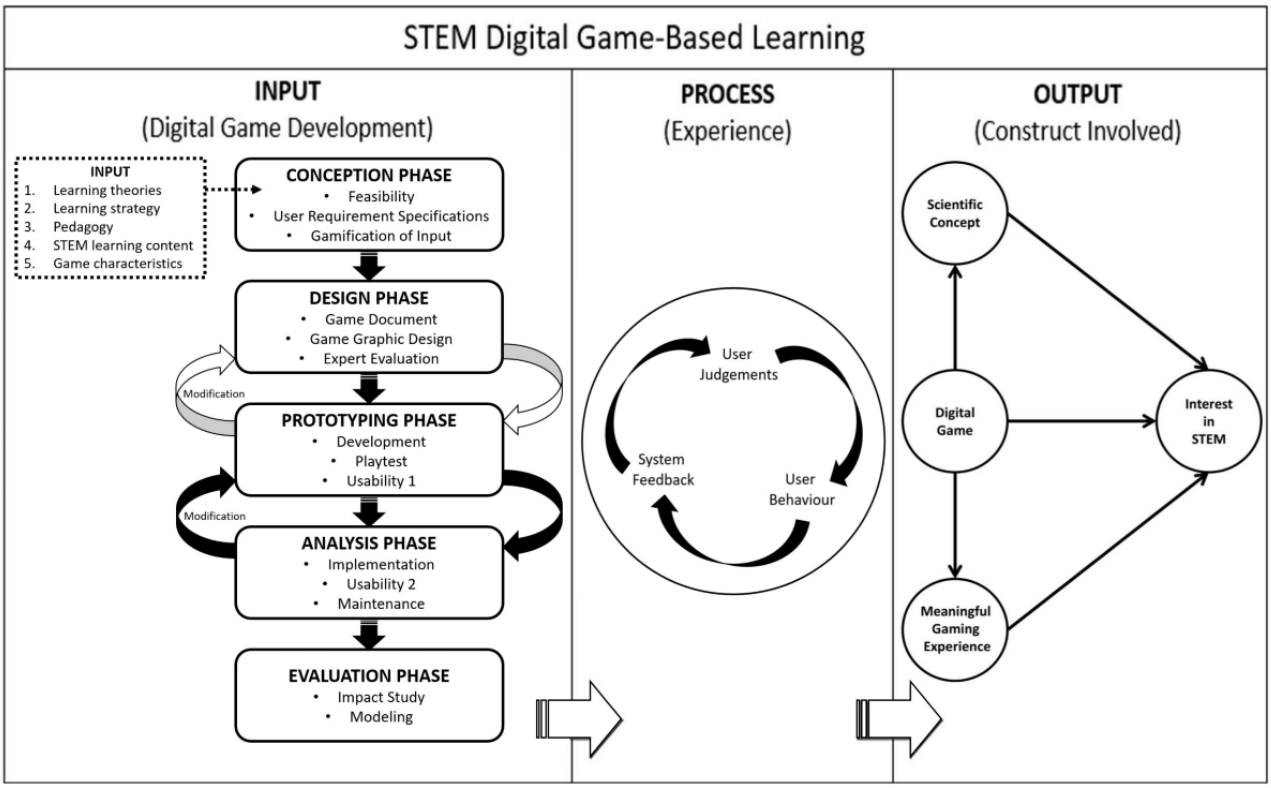
Variables	Findings	References
Digital games → scientific concepts	Digital games are potential tools for teaching STEM[a] concepts via a specific topic. Knowledge gains underlie the gameplay. The integration of scientific content acts as a vehicle for gameplay. Most players exhibit higher knowledge gain, as a STEM learning outcome, using digital games. The target group, with the most potential for STEM digital games, is children.	[9,31,32,54-61]
Digital games → meaningful gaming experience	Playing digital games provides a gaming experience. As part of edutainment, good digital games create meaningful experiences for players. Players with a higher level of game experience and feedback engage in a more in-depth reflective process. Players will feel the fun element and have strong connections with the character played inside the game world.	[62-69]
Digital games → interest in STEM	Digital media have a significant relationship with an interest in STEM. Due to better gameplay, digital games might therefore be successful in increasing students' interest in STEM. Studies show that using digital games may lead to a positive attitude toward STEM subjects and a STEM career.	[2,11,55,70-72]
Scientific concepts → interest in STEM	Knowledge gained by providing authentic scientific inquiry will trigger an interest in STEM. Students who play video games will engage in a STEM subject related to their lives.	[55,73-75]
Meaningful gaming experience → interest in STEM	Providing digital games will serve to enhance individual curiosity. Once attracted to the game, the playing process will increase the level of interest in the content gained. Because a meaningful gaming experience is a product of playing digital games, studies show that this factor is significantly related to an interest in STEM.	[76-79]

[a]STEM: science, technology, engineering, and mathematics.

To measure the relationship between the 4 identified variables (digital games, scientific concepts, meaningful gaming experience, and interest in STEM) simultaneously, these variables need to be defined and measured individually in the form of a measurement model. Structural equation modeling helps to measure and fit this hypothetical model with data in one go. This outcome at this stage is best achieved using smart-PLS software (SmartPLS GmbH) since it is at an exploratory stage.

The proposed STEM digital game-based learning framework is that of a linear input-process-output structure (Figure 8). This framework can be employed by any game designer, developer, or even teacher to develop a good STEM digital game with the objective of increasing STEM interest among learners. Young people with a low interest in STEM will therefore benefit from playing a STEM digital game.

Figure 8. Proposed framework on STEM digital game-based learning.



Conclusion

This paper proposes a framework for STEM digital game-based learning based on the current and vast body of gaming technology and a universal design for learning instructional design and agile development processes. Following the proposed model will enable the product to compete with what is available in a market where most developers are competing to produce digital games. However, the game needs to undergo a systematic development process as proposed. This paper thus outlined the specifications needed for a digital game to obtain the desired outcomes.

Acknowledgments

We would like to convey our utmost appreciation and thanks to all those who supported our study, especially the STEM Enculturation Research Centre (Faculty of Education, *Universiti Kebangsaan Malaysia*) for grants *Dana Pecutan Penerbitan* (GUP-2020-020 and GG-2019-046). Many thanks to all researchers under the project and the Personalized Education Research Group for financial, intellectual, spiritual, and moral support.

Conflicts of Interest

None declared.

References

1. Kukulska-Hulme A, Bossu C, Coughlan T, Ferguson R, FitzGerald E, Gaved M, et al. Innovating Pedagogy 2021: Open University Innovation Report 9. Milton Keynes: The Open University; 2021:1-51.
2. Ball C, Huang K, Cotten SR, Rikard RV. Gaming the SySTEM: the relationship between video games and the digital and STEM divides. Games Cult 2018 Nov 29;15(5):501-528. [doi: 10.1177/1555412018812513]
3. Mohtar LE, Halim L, Abd Rahman N, Maat SM, Iksan ZH, Osman K. A model of interest in stem careers among secondary school students. J Baltic Sci Education 2019 Jun 10;18(3):404-416. [doi: 10.33225/jbse/19.18.404]
4. Zendler A, Greiner H. The effect of two instructional methods on learning outcome in chemistry education: the experiment method and computer simulation. Education Chem Engineers 2020 Jan;30:9-19. [doi: 10.1016/j.ece.2019.09.001]
5. Yousof A, Abousamra N. The future of digital game-based learning (DGBL) in shaping intercultural communication competency. In: Lum HC, editor. Critical Issues Impacting Science, Technology, Society (STS), and Our Future. Hershey, PA: IGI Global; 2019:81-110.
6. Mat Sin N, Talib O, Norishah TP. Merging of game principles and learning strategy using apps for science subjects to enhance student interest and understanding. Jurnal Teknologi 2013 Aug 15;63(2):7-12. [doi: 10.11113/jt.v63.1998]

7. Freeman B, Marginson S, Tytler R. The Age of STEM: Educational Policy and Practice Across the World in Science, Technology, Engineering and Mathematics. United Kingdom: Routledge; Nov 3, 2014.
8. Hyatt KJ, Barron JL, Noakes MA. Video gaming for STEM education. In: Yang HH, Wang S, editors. Cases on E-Learning Management: Development and Implementation. Hershey, PA: IGI Global; 2013:103-117.
9. Wu Y, Anderson OR. Technology-enhanced STEM (science, technology, engineering, and mathematics) education. J Comput Educ 2015 Jul 24;2(3):245-249. [doi: 10.1007/s40692-015-0041-2]
10. Lay A, Osman K. Developing 21st century chemistry learning through designing digital games. J Educ Environ Sci Health 2018;4(1):81-92. [doi: 10.21891/jeseh.387499]
11. Fisher C. Designing Games for Children: Developmental, Usability, and Design Considerations for Making Games for Kids. New York: Focal Press; 2015.
12. Acquah EO, Katz HT. Digital game-based L2 learning outcomes for primary through high-school students: a systematic literature review. Comput Education 2020 Jan;143:103667-103619. [doi: 10.1016/j.compedu.2019.103667]
13. Garris R, Ahlers R, Driskell JE. Games, motivation, and learning: a research and practice model. Simulation Gaming 2016 Aug 18;33(4):441-467. [doi: 10.1177/1046878102238607]
14. Richards J, Smith C. 100 Top Games Apps Made Easy. London: Flame Tree Publishing; 2013.
15. Aleem S, Capretz LF, Ahmed F. Game development software engineering process life cycle: a systematic review. J Softw Eng Res Dev 2016 Nov 9;4(1):1-30. [doi: 10.1186/s40411-016-0032-7]
16. Din R. Foreword from the chief editor: the inaugural issue of JPL. J Personalized Learning 2018 May 1;1(1):i-iii [FREE Full text]
17. Din R. Notes from the chief editor: on designing personalized learning. J Personalized Learning 2016;2(1):i-iii [FREE Full text]
18. Din R. Model Pembangunan Sistem-Pembinaan Sistem Persidangan Berkomputer: SidangKom. Malaysian Academic Library Union Catalog.: Universiti Kebangsaan Malaysia; 2001. URL: http://malcat.uum.edu.my/kip/Record/ukm.vtls000327247 [accessed 2021-01-26]
19. Din R. Pembinaan dan Permodelan Sistem Pengajaran. Bangi: Universiti Kebangsaan Malaysia; 2014.
20. Din R. Development and validation of an integrated meaningful hybrid e-training (I-MeT) for computer science: theoretical and empirical based design and development approach. ResearchGate. Bangi: Universiti Kebangsaan Malaysia; 2010. URL: tinyurl.com/32vuz9mo
21. Din R. PLS-SEM in the final phase of product and instrument development using UDIN model. 2019 Presented at: 11th International Conference of Numerical Analysis in Engineering; August 20-22; Bandar Baru Bangi.
22. Din R. Asas Pendidikan dan Kejurulatihan ICT: Integrasi Teori, Media, Teknologi dan Reka Bentuk Pembelajaran. Bangi: Penerbit UKM; 2017.
23. Din R, Othman N, Nasir M, Murat N, Alias M. 2018 Presented at: Learning Carnival and Conference; July 16; Melaka.
24. Din R, Othman N, Murat N, Alias M, Nasir M. Formative to summative in one go. 2017 Presented at: Holistic Learning Conference (Rahole); November 20-21; Kuala Lumpur.
25. Din R, Othman N, Murat N, Alias M, Nasir M. Reflection in a blog: scaffolding to formative and summative assessment. 2017 Presented at: Holistic Learning Conference (Rahole); November 20-21; Kuala Lumpur.
26. Marcelo LE, Gabriela TP, Ivana LL, Thiago TM. Mobile game-based learning in STEM subjects. In: Khosrow-Pour M, editor. Encyclopedia of Information Science and Technology, Fourth Edition. Hershey, PA: IGI Global; 2018:6376-6387.
27. Mohdjamel F, Ahmad JN, Ali MN. The needs analysis in game-based STEM module development for KSSM science teachers. Int J Recent Technol Engineering 2019 Sep 30;8(3):6622-6628. [doi: 10.35940/ijrte.c5655.098319]
28. Cunningham L, Murphy O. Embracing the universal design for learning framework in digital game based learning - a set of game design principles. Stud Health Technol Inform 2018;256:409-420. [Medline: 30371502]
29. Masrop N, Zin N. Model Reka Bentuk dan Pembangunan Perisian Pembelajaran Berasaskan Permainan Digital (PBPD). In: Mat Zain NA, Sahari@Ashaari N, editors. Reka Bentuk dan Kebolehgunaan Sistem Multimedia. Bangi: Penerbit UKM; 2017:82-95.
30. Shi Y, Shih J. Game factors and game-based learning design model. Int J Comput Games Technol 2015;2015:1-11. [doi: 10.1155/2015/549684]
31. Ross AM, Fitzgerald ME, Rhodes DH. Game-based learning for systems engineering concepts. Procedia Comput Sci 2014;28:430-440. [doi: 10.1016/j.procs.2014.03.053]
32. Shute VJ, Ke F. Games, learning, and assessment. In: Ifenthaler D, Eseryel D, Ge X, editors. Assessment in Game-Based Learning: Foundations, Innovations, and Perspectives. New York: Springer Science + Business Media; 2012:43-58.
33. Norizan MD, Nor Azan MZ. Reka Bentuk Permainan Digital Latihan Menulis Huruf Jawi. In: Nor Azan MZ, Noraidah SA, editors. Reka Bentuk dan Kebolehgunaan Sistem Multimedia. Bangi: Unuversiti Kebangsaan Malaysia; 2017:63-81.
34. Kirkley SE, Tomblin S, Kirkley J. Instructional Design Authoring Support for The Development of Serious Games and Mixed Reality Training. 2005 Presented at: Interservice/Industry Training, Simulation and Education Conference (I/ITSEC); December 6-9; Orlando, Florida.
35. Jurva H. Could universal design help more students learn from games? Games and learning. 2015 Aug 13. URL: https://www.gamesandlearning.org/2015/08/13/could-universal-design-help-more-students-learn-from-games/ [accessed 2020-03-28]

36. Anderson F. How true universal design could revolutionize game development. Venture Beat. 2010 Dec 2. URL: https://venturebeat.com/community/2010/12/02/how-true-universal-design-could-revolutionize-game-development/ [accessed 2020-03-28]
37. Tan WH. Gamifikasi dalam Pendidikan: Pembelajaran Berasaskan Permainan. Tanjong Malim: Penerbit Universiti Pendidikan Sultan Idris; 2016.
38. Rapp A, Hopfgartner F, Hamari J, Linehan C, Cena F. Strengthening gamification studies: current trends and future opportunities of gamification research. Int J Hum Comput Stud 2019 Jul;127:1-6. [doi: 10.1016/j.ijhcs.2018.11.007]
39. Ak O. A game scale to evaluate educational computer games. Procedia - Soc Behav Sci 2012;46:2477-2481. [doi: 10.1016/j.sbspro.2012.05.506]
40. Mat Zin NA, Yue WS. Design and evaluation of history digital game based learning (DGBL) software. J Next Generation Inf Technol 2013 Jun 30;4(4):9-24. [doi: 10.4156/jnit.vol4.issue4.2]
41. Pulsipher L. Game Design-How to Create Video and Tabletop Games, Start to Finish. North Carolina: McFarland & Company Publisher; 2012.
42. de Byl P. Holistic Mobile Game Development with Unity. New York: Routledge; 2014.
43. Mildner P, Stamer N, Effelsberg W. From game characteristics to effective learning games. 2015 Presented at: Serious Games: First Joint International Conference; June 3-4; Huddersfield, UK p. 51-62. [doi: 10.1007/978-3-319-19126-3_5]
44. Lohr L, Eikleberry C. Learner-centered usability. tools for creating a J learner-friendly instructional environment. Performance Improvement 2001 Apr;40(4):24-27. [doi: 10.1002/pfi.4140400406]
45. Nielson J. Usability Engineering. University of Minnesota, US: Academic Press; 1993.
46. Din R. Notes from the chief editor: on universal design and agile development. Journal of Personalized Learning 2020;3(1):1-7.
47. Din R, Hamdan A, Manaf SZA. Edutechnovation day to measure learning outcome for acceptable hybrid meaningful mobile learning environment. 2015 Presented at: National University Carnival on e-Learning; December 14-15; Tanjung Malim, Perak.
48. Azizul S, Din R. Teaching and learning geometry using Geogebra software via MOOC. J Personalized Learning 2016;2(1):39-50.
49. Bataniah R, Din R. Hybrid personalized Arabic language learning. J Personalized Learning 2016;2(1):57-71.
50. Manaf S, Din R. Hamdan A Pembelajaran Peribadi Melalui Teknologi ICT Abad ke-21 ke Arah Kemahiran Boleh Pindah. J Personalized Learning 2015;1(1):57-68.
51. Hamdan A, Din R, Manaf SZA. Personalized learning environment: integration of web technology 2.0 in achieving meaningful learning. J Personalized Learning 2015;1(1):13-26.
52. Breiner J, Johnson C, Harkness S, Koehler C. What is STEM? a discussion about conceptions of STEM in education and partnerships. Sch Sci Math 2012;112(1):3-11. [doi: 10.1111/j.1949-8594.2011.00109.x]
53. Bryan L, Moore T, Johnson C, Roehrig G. Integrated STEM education. In: Johnson CC, Peters-Burton EE, Moore TJ, editors. STEM Road Map: A Framework for Integrated STEM Education. New York: Routledge; 2016:23-37.
54. Herodotou C. Mobile games and science learning: a comparative study of 4 and 5 years old playing the game Angry Birds. Br J Educ Technol 2017 Feb 13;49(1):6-16. [doi: 10.1111/bjet.12546]
55. Raffety C, Prawat T, Richter J, Hamilton RF, Schelvan M, Jones P, et al. Developing Serious games to improve learning and increase interest in STEM careers for middle school students: the mice of riddle place. 2016 Presented at: Immersive Learning Research Network: Second International Conference; June 27 – July 1; Santa Barbara, CA, USA p. 132-143. [doi: 10.1007/978-3-319-41769-1_11]
56. Liu EZF, Chen P. The effect of game-based learning on students' learning performance in science learning – a case of "conveyance go". Procedia Soc Behav Sci 2013 Nov;103:1044-1051. [doi: 10.1016/j.sbspro.2013.10.430]
57. Pedersen MK, Svenningsen A, Dohn NB, Lieberoth A, Sherson J. DiffGame: game-based mathematics learning for physics. Procedia Soc Behav Sci 2016 Jul;228:316-322. [doi: 10.1016/j.sbspro.2016.07.047]
58. Sherry JL. Formative research for STEM educational games: lessons from the Children's Television Workshop. In: Blades M, Blumberg FC, Oates C, editors. Children and Digital Games. USA: Hogrefe Publishing; 2013:90-97.
59. Clark DB, Sengupta P, Brady CE, Martinez-Garza MM, Killingsworth SS. Disciplinary integration of digital games for science learning. Int J STEM Educ 2015 Feb 11;2(1):2-21. [doi: 10.1186/s40594-014-0014-4]
60. Kiili K, Ketamo H. Evaluating cognitive and affective outcomes of a digital game-based math test. IEEE Trans Learning Technol 2018 Apr 1;11(2):255-263. [doi: 10.1109/tlt.2017.2687458]
61. Lester JC, Spires HA, Nietfeld JL, Minogue J, Mott BW, Lobene EV. Designing game-based learning environments for elementary science education: a narrative-centered learning perspective. Inf Sci 2014 Apr;264:4-18. [doi: 10.1016/j.ins.2013.09.005]
62. Hou H. Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. Comput Human Behav 2015 Jul;48:424-435. [doi: 10.1016/j.chb.2015.02.010]
63. Kiili K, de Freitas S, Arnab S, Lainema T. The design principles for flow experience in educational games. Procedia Comput Sci 2012;15:78-91. [doi: 10.1016/j.procs.2012.10.060]

64. Calvillo-Gámez EH, Cairns P, Cox AL. Assessing the core elements of the gaming experience. In: Bernhaupt R, editor. Evaluating User Experience in Games, Human-Computer Interaction Series. London: Springer; 2010:47-71.
65. Rogers R, Woolley J, Sherrick B, Bowman ND, Oliver MB. Fun versus meaningful video game experiences: a qualitative analysis of user responses. Comput Game J 2016 Nov 25;6(1-2):63-79. [doi: 10.1007/s40869-016-0029-9]
66. Mills DJ, Milyavskaya M, Mettler J, Heath NL. Exploring the pull and push underlying problem video game use: a self-determination theory approach. Personality and Individual Differences 2018 Dec;135:176-181. [doi: 10.1016/j.paid.2018.07.007]
67. Sajjadi P, Hoffmann L, Cimiano P, Kopp S. A personality-based emotional model for embodied conversational agents: effects on perceived social presence and game experience of users. Entertainment Computing 2019 Dec;32:100313-100310. [doi: 10.1016/j.entcom.2019.100313]
68. Tammy Lin J, Bowman N, Lin S, Chen Y. Setting the digital stage: defining game streaming as an entertainment experience. Entertainment Computing 2019 Aug;31:1-8. [doi: 10.1016/j.entcom.2019.100309]
69. Oliver MB, Bowman ND, Woolley JK, Rogers R, Sherrick BI, Chung M. Video games as meaningful entertainment experiences. Psychology of Popular Media Culture 2016 Oct;5(4):390-405. [doi: 10.1037/ppm0000066]
70. Bonner D, Dorneich M. Developing game-based learning requirements to increase female middle school students interest in computer science. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2016 Presented at: Human Factors and Ergonomics Society Annual Meeting; September 19-23; Washington, DC p. 380-384. [doi: 10.1177/1541931213601086]
71. Halim L, Rahman N, Wahab N, Mohtar L. Factors influencing interest in STEM careers: an exploratory factor analysis. Asia-Pacific Forum on Science Learning and Teaching 2018;19(2):1-34.
72. Krapp A. Interest, motivation and learning: an educational-psychological perspective. Eur J Psychol Educ 1999 Mar;14(1):23-40. [doi: 10.1007/bf03173109]
73. Martin-Hansen L. Examining ways to meaningfully support students in STEM. Int J STEM Educ 2018 Dec 20;5(1):53. [doi: 10.1186/s40594-018-0150-3]
74. Guzey SS, Ring-Whalen EA, Harwell M, Peralta Y. Life STEM: a case study of life science learning through engineering design. Int J Sci Math Educ 2017 Oct 6;17(1):23-42. [doi: 10.1007/s10763-017-9860-0]
75. Beier ME, Rittmayer AD. Motivational factors in STEM: interest and self concept. In: Bogue B, Cady E, editors. Applying Research to Practice (ARP) Resources. Canada: ARP Resources; 2008:1-10.
76. Roberts T, Jackson C, Mohr-Schroeder MJ, Bush SB, Maiorca C, Cavalcanti M, et al. Students' perceptions of STEM learning after participating in a summer informal learning experience. Int J STEM Educ 2018 Sep 21;5(1):1-14. [doi: 10.1186/s40594-018-0133-4]
77. Mohd Shahali EH, Halim L, Rasul MS, Osman K, Zulkifeli MA. STEM learning through engineering design: impact on middle secondary students' interest towards STEM. EURASIA J Math Sci Tech Ed 2017;13(5):1189-1121. [doi: 10.12973/eurasia.2017.00667a]
78. Baxter J, Collins J. STEM education: creating meaningful experiences with interaction design. In: Reidsma D, editor. Intelligent Technologies for Interactive Entertainment. USA: Springer International Publishing; 2014:109-112.
79. Meaningful experiences in STEM learning make a difference!. Project Exploration. 2017. URL: https://projectexploration.org/2017/12/21/meaning-experiences-stem-learning-make-difference/ [accessed 2020-08-10]

Abbreviations

STEM: science, technology, engineering, and mathematics

Edited by G Eysenbach; submitted 21.05.20; peer-reviewed by C Gabarre, P Stourac; comments to author 15.06.20; revised version received 09.10.20; accepted 24.10.20; published 19.02.21.

Please cite as:
Ishak SA, Din R, Hasran UA
Defining Digital Game-Based Learning for Science, Technology, Engineering, and Mathematics: A New Perspective on Design and Developmental Research
J Med Internet Res 2021;23(2):e20537
URL: http://www.jmir.org/2021/2/e20537/
doi:10.2196/20537
PMID:33605885

Viewpoint

The Computer Will See You Now: Overcoming Barriers to Adoption of Computer-Assisted History Taking (CAHT) in Primary Care

Pier Spinazze[1], MD, MSc, MBA; Jiska Aardoom[2], PhD; Niels Chavannes[2], MD, PhD; Marise Kasteleyn[2], PhD

[1]Global Digital Health Unit, Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

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

Corresponding Author:
Pier Spinazze, MD, MSc, MBA
Global Digital Health Unit
Department of Primary Care and Public Health, School of Public Health
Imperial College London
Dunstan's Road
London, W6 8RP
United Kingdom
Phone: 44 7426965665
Email: p.spinazze@imperial.ac.uk

Abstract

Patient health information is increasingly collected through multiple modalities, including electronic health records, wearables, and connected devices. Computer-assisted history taking could provide an additional channel to collect highly relevant, comprehensive, and accurate patient information while reducing the burden on clinicians and face-to-face consultation time. Considering restrictions to consultation time and the associated negative health outcomes, patient-provided health data outside of consultation can prove invaluable in health care delivery. Over the years, research has highlighted the numerous benefits of computer-assisted history taking; however, the limitations have proved an obstacle to adoption. In this viewpoint, we review these limitations under 4 main categories (accessibility, affordability, accuracy, and acceptability) and discuss how advances in technology, computing power, and ubiquity of personal devices offer solutions to overcoming these.

(J Med Internet Res 2021;23(2):e19306) doi:10.2196/19306

KEYWORDS

computer-assisted history taking; history taking; clinical consultation; digital health; electronic health record; patient-provided health information

Introduction

Background

Computer-assisted history taking (CAHT) was first explored in medicine in the 1960s as a tool to aid clinicians in gathering data from patients in order to support diagnostic and treatment decisions. Over the years, research has highlighted the numerous benefits of CAHT; however, the limitations have proved an obstacle to adoption. We explore how advances and innovation in digital technology today may offer new solutions to overcoming these barriers.

The Importance of History Taking in Clinical Care

Collecting an appropriate and comprehensive medical history from patients is a fundamental process in clinical medicine. Research has shown that history taking alone is sufficient to make a diagnosis in 75% of patient encounters, before further reducing the number of clinical differentials through physical examination and additional tests [1]. However, the amount of time available to acquire an appropriate patient history is decreasing. The increasing demand and administrative burden on healthcare services have resulted in physician-patient contact time becoming shorter. A study in the United States highlighted that physicians spent 27% of their total time on direct clinical face time with patients and 49.2% of their time on the electronic health record (EHR) and deskwork [2]. During in-person consultation, on average, only 52.9% of physicians' time was spent on direct clinical face time and 37.0% on EHR and deskwork [2]. The consultation length is directly associated with better health outcomes, fewer prescriptions, and better recognition of long-term and psychosocial problems [3,4]. On the other hand, a shorter consultation time has been associated with overuse of antibiotics, polypharmacy, and poor

communication with patients [5,6]. Whether this is attributable to diagnostic uncertainty or fear of medicolegal repercussions remains to be evaluated.

CAHT vs In-Person Interviews

Considering restrictions to consultation time and the associated negative consequences, patient-provided health data outside of consultation can prove invaluable in health care delivery. Several health bodies, including the Centers for Disease Control and Prevention, advocate for the serial collection of health surveys in routine clinical care [7]. These have traditionally been completed with pen and paper and are commonly not collected due to numerous barriers including difficulties in logistics of acquisition, distribution, and collection of paper forms; difficulties in understanding and completing surveys by patients; the potential disruption of clinic workflow; difficulties in interpreting results; lack of perceived clinical relevance; and cost (materials, manpower, and distribution) [8].

A CAHT system (CAHTS) is a digital tool that aids clinicians in gathering data from patients through health surveys to inform a diagnosis or treatment plan [9]. The benefits of CAHTS are evident in the potential time saving in terms of acquiring the patient history outside of consultation, reducing the administrative burden of entering this information, increasing patient face-to-face time, and leveraging these data through medical records using machine learning algorithms for decision support. Patients have reported high satisfaction from helping their physician through the completion of interactive computerized interviews [10]. CAHT is an effective strategy to empower patients to be active in their own care (ie, patient engagement) [11]. Increased patient involvement results in improved participation in personal care, compliance with medication, adherence to recommended treatment, and monitoring of prescriptions and doses (for a complete overview of potential advantages, see Textbox 1).

Textbox 1. Potential advantages of computer-assisted history taking (CAHT) [9,12-16].

- Enables history taking prior to or post consultation
- Enables completion at the patient's pace without feeling rushed
- High compliance rate (for patients)
- Time efficient (for both patient and clinician)
- Collects complete and accurate patient data
- Provides legible summaries
- Less likely to have falsified data
- Patients reveal more sensitive "private" and social information
- Patients are better prepared for the medical interview
- Reduces staff labor costs
- Reduces diagnostic error
- Enables remote completion
- Allows health care professionals to make additional entries or changes
- Allows for seamless integration into the patient electronic health record (EHR)
- Can incorporate artificial intelligence and decision support systems
- Can prompt educational messages or modules

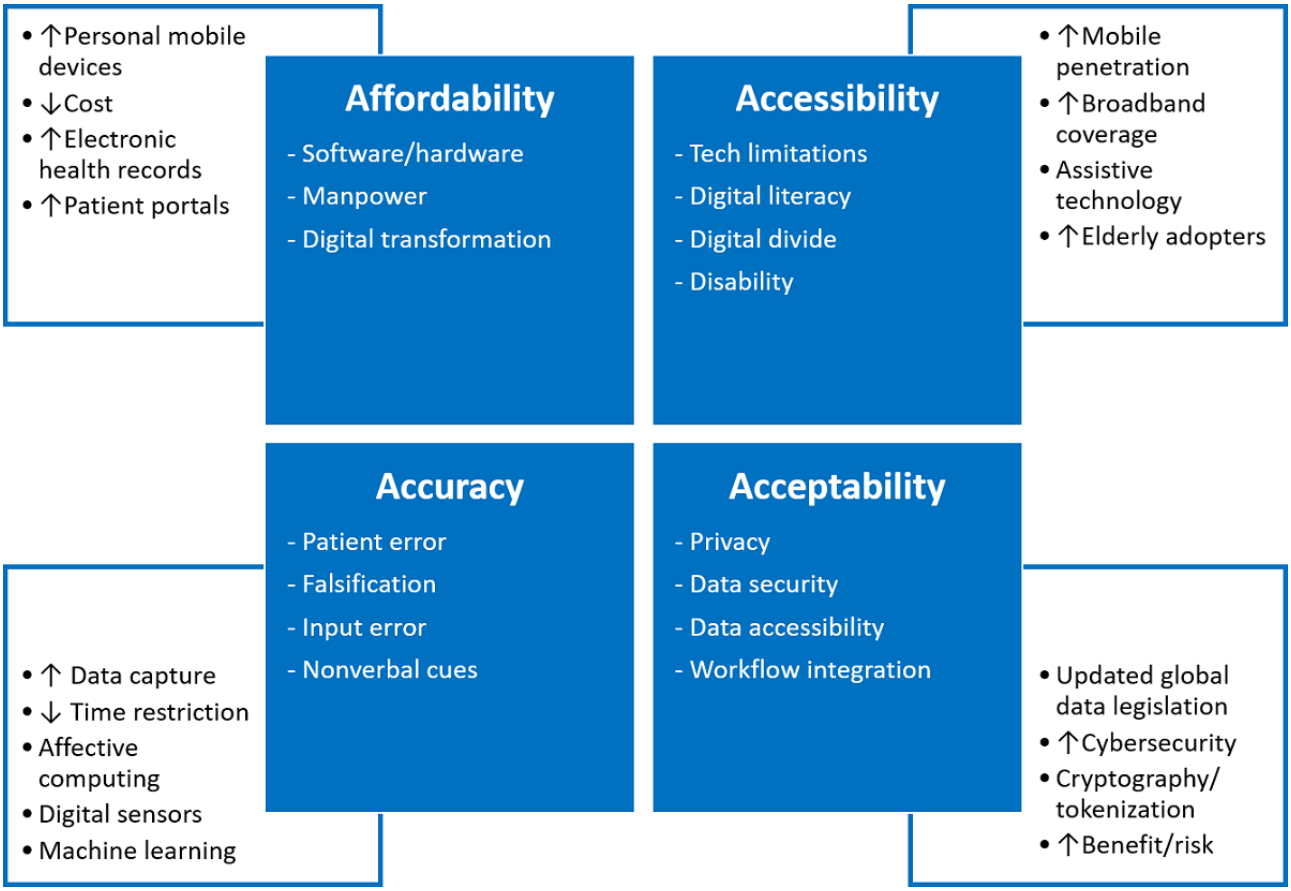
CAHT can be done remotely through mobile devices at a time and place convenient to the patient. Furthermore, CAHT can improve the comprehensiveness of history taking by standardizing algorithms and extending the level of questioning through branching logic based on participant responses. Research has shown that clinicians are ineffective at acquiring comprehensive patient histories; they tend to miss 50% of psychosocial and psychiatric problems and do not elicit 54% of patient health problems [9].

Perhaps one of the greatest advantages of CAHTS is providing a shared knowledge base of clinical presentations and outcomes that is readily accessible to all providers. Decision making, especially in medicine, is complex and multifaceted. The accuracy of decisions is directly linked to whether the relevant information is readily available from memory [17]. One doctor's decision process is directly linked to personal experience and whether that knowledge is readily retrievable (ie, comes to mind or is accessible at the time from sources such as textbooks, online libraries). In contrast, a digitized database allows for combined case histories, including multiple different variables (eg, lab results, imaging) to be stored, processed, and rapidly accessed to support decision making. This could greatly facilitate diagnostic accuracy, especially in rarer diagnostic cases.

Overcoming Limitations to CAHT

It is clear that there are many advantages to CAHT; however, its limitations could explain the slow rates of adoption of patient-provided health information in general [18]. We have grouped these limitations under 4 themes identified through literature reviews: accessibility, affordability, accuracy, and acceptability. We subsequently discuss how technological advances offer solutions to overcoming these (see Figure 1).

Figure 1. Overview of barriers to implementation of computer-assisted history taking systems (CAHTS) and changing landscape and technological solutions.



Accessibility

During the inception of CAHT in 1960, technology represented a potential barrier to implementation due to accessibility to computers. However, since then, innovation in mobile computing devices has seen a proliferation in personal devices from PDAs, followed by smartphones and tablets. Mobile devices have become ubiquitous accessories with more mobile devices on the planet now than people [19]. The digital divide between low- and middle-income countries and high-income countries is progressively narrowing, as mobile phone penetration rates and mobile broadband and internet access continue to expand [20]. The number of smartphone users globally has just surpassed 3 billion [21], with penetration rates that range from around 24% in India to 90% in high-income countries [20]. Hence, with computing power, communications, and internet now being untethered to physical, on-premise computers, access is no longer a limiting factor but rather facilitates the ability to deliver CAHT.

The early adopters of technology, including mobile phones, are commonly the young, tech-savvy, and able-bodied persons, previously limiting those who are older, digitally naïve, and disabled (hearing or visually disabled) from utilizing them. However, the elderly are increasingly becoming avid adopters of technology. In the United States, the fastest growing segment of smartphone adopters are aged 44-75 years old with a compound annual growth rate of close to 8% from 2015-2017 [22]. Mobile devices are being designed with assistive technology to further facilitate use. In mobile phones, for example, there are now a variety of accessibility options available, including text-to-speech output, screen magnification, audio amplifiers, hearing aid compatibility, and hands-free operation [23], facilitating use by those with visual or hearing disabilities. Assistive technology is also leveraging embedded systems and wearables. A “smart glove” is one such example, which recognizes basic hand gestures and converts them into speech or text [24]. Wearables allow for the passive collection of patient data, both physiological and pathological information, continuously and in real time, directly into EHRs. This reduces the barrier to acquiring data actively and can be used in health and safety monitoring, chronic disease management, disease diagnosis and treatment, and rehabilitation [25]. These data points could further augment CAHT; however, further research would need to be done to evaluate this benefit.

Beyond assistive technology, CAHT can be delivered in multiple formats including visually rich media, audible questionnaires, and multiple languages, tailoring the content to the targeted participant group. In terms of user accessibility, therefore, it is increasingly becoming ubiquitous and accessible by all.

The great strides taken to digitize health data and integrate EHRs have resulted in a highly fragmented system wherein health data are distributed in silos across the continuum of care with limited accessibility between providers and systems. Lack of interoperability and access to data therefore limit the potential benefits of CAHT. As in other industries, the barriers to interoperability are generally not technological but cultural and require the close coordination and collaboration of various

stakeholders, including patients, providers, software vendors, legislators, and information technology professionals [26]. Over the last few years, there have been a number of improvements in policy, alignment of incentives, and the wider adoption of data-sharing protocols and infrastructure to help mitigate these accessibility issues. A number of health information technology solutions, such as health information exchanges, have propagated the collaboration of health systems to integrate across many different digital silos [27].

Affordability

Although CAHTS are viewed as cost-effective compared to regular standard of care, few studies have rigorously assessed the cost-benefit. Of those that have, most of the data available are related to changes in utilization of health care services due to health information technology [28]. There is a number of cost factors that can be taken into account, including equipment, time, manpower, and other costs. In practice, a CAHTS can be integrated into an EHR to allow for a data record and to leverage data insights accumulated from multiple sources, including lab results, imaging, and doctors' records. In isolation, the benefits of a CAHT could not be fully realized. Hence, equipment costs when considering deploying a health information system or EHR can be significantly high. These costs can be categorized as system costs and induced costs [29]. System costs include the costs of software and hardware, training, implementation, ongoing maintenance, and support. Induced costs are related to the temporary productivity loss during EHR implementation. Several studies estimate the cost of purchasing and installing an EHR ranging from US $15,000 to US $70,000 per provider [30] with subsequent yearly maintenance costs. Previously, these costs may have presented a strong financial barrier to implementing CAHTS; however, EHR adoption rates have vastly increased. From 2008 till 2017, office-based physician adoption of EHRs in the United States has more than doubled, from 42% to 86% [31], with nearly universal adoption in the United Kingdom, Netherlands, Australia, and New Zealand [32]. Hence, considering the widespread prevalence of EHRs and initial sunk costs, the additional cost of implementing CAHTS is very low. Three of the largest EHR vendors (Epic, Allscripts, and Cerner) [33] all offer the capability to send questionnaires to patients and to receive structured data directly into the medical record. Although cost barriers to implementation have been greatly reduced over the past decades, a comprehensive analysis would need to be done to evaluate actual costs.

The cost of accessing CAHTS and health care systems outside of premises is also borne by the patients. The biggest factor limiting online access is cost, with the internet and mobile phones still not affordable for many around the world. Two and half billion people live in countries where the cost of the cheapest available smartphone is >25% of the average monthly income [34]. A number of organizations are seeking to improve these numbers, including the Web Foundation, the United Nations, and other national governments, which has been further necessitated by the COVID-19 pandemic. A part of the United Nation's Sustainable Development Goals is to achieve "universal access" in the least developed countries by the end of 2020. The Web Foundation defines "universal access" as 85% of a country's population [35].

Accuracy

Previously, there was some skepticism regarding the accuracy of CAHT in the sense that patients may not provide accurate information through a digital survey as opposed to in-person consultation with a physician. Postulated reasoning for this includes patient's failure to read questions and answers carefully, misunderstood questions, mistaken selection of answers, failure to comprehend prior diagnoses, and intentional entry of false information [36]. The accuracy of patient's answers in terms of mistaken entries, not deliberate falsification of inputs, is a significant issue for both computerized and physician history taking [36]. Provider errors beyond communication barriers can occur in patient notetaking, including incomplete patient notes and illegible handwriting, as well as errors in data input or transcription into EHRs. An analysis of medical malpractice cases found that incorrect information (eg, faulty data entry) was the top EHR-related contributing factor, contributing to 20% of reviewed cases [37]. Essentially, human entry error will always be a factor, both on the patient and provider side; however, false information can potentially be reduced through CAHTS. There is substantial evidence to support that direct reporting of symptoms by patients through CAHTS more accurately reflects their health status than through clinician elicitation [38]. Face-to-face interview methods typically result in reporting of lower rates of socially sensitive risk behaviors compared to self-administered questionnaires, attributable to social desirability bias [39]. For example, patients are more likely to report sensitive information including intimate partner violence, elective abortions, and high-risk behavior such as smoking status in computer interviews [40]. The nature of the doctor-patient relationship and demographic (eg, different cultural norms, age, and gender) may also influence patient reporting [41]. Another source of error includes unintentional reporting error due to poor recall and situational or time pressure to respond (ie, the patient feels they are on the spot or the doctor is rushing them). The benefit to CAHT is that it is not confined to consultation time, therefore is not time restricted, and the questions may be algorithm-based allowing for a structured, comprehensive collection of data rather than ad hoc questioning dependent on the provider's thought process.

CAHTS are inherently limited by the inability to record nonverbal communication. Computers are "unable to detect nonverbal behavior, for example, sense a patient's mood which might easily be picked up in a consultation" [4]. However, technology has since evolved to be able to capture various nonverbal behavioral cues like facial expressions, vocalizations, postures, gestures, and appearance. There have been significant advances in the field of affective computing, which is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects [42]. It is an interdisciplinary field that leverages the crosspollination of artificial intelligence (including speech processing, computer vision, and machine learning) and human sciences (psychology, anthropology, and sociology). By utilizing digital sensor data such as video or sound recordings, behavioral patterns can be evaluated. For example, microexpressions (involuntary, brief

facial movements) play an essential part in understanding nonverbal communication and deceit detection [43]. Due to the nature of microexpressions being extremely brief and subtle, it is difficult to perceive with the naked eye. Advances in computer algorithms and video acquisition technology are rendering machine analysis of facial microexpressions an increasing possibility [44]. Facial expressions can be used in combination with other modalities including head and hand movements to detect deception [45].

Although there is still a long road to go towards accurate real-time assessment, we can now leverage several modalities to detect nonverbal cues while completing CAHT. Natural language processing (NLP) may offer a more readily available solution to deceit detection through text analysis. In recent years, with the explosive popularity of social media and subsequent exposure of fake news, increasing attention has been put on lie detection using artificial intelligence and deep learning techniques. Beyond falsification of information, NLP can analyze sentiment, which includes the words and symbols used in text to indicate positive and negative opinions, and emotions. Sentiment analysis has been well studied in health care, including its relation to outcomes (eg, greater positive sentiment within discharge summaries) associated with significantly decreased risk of readmission [46]. Through technology, we may therefore be able to predict mental state or falsification of information with potentially greater accuracy than by clinicians.

Acceptability

Despite the benefits of CAHTS, a key barrier to adoption could be acceptability to both patients and providers. Previous challenges included acceptability in terms of technology and associated challenges with its use [47]. This may have been the case in the past with unreliable platforms and outdated devices (eg, PDAs). However, studies have shown that patients report a high level of satisfaction with CAHT, with the majority (69%) believing that their medical care is enhanced by CAHTS [36]. Youth find computerized questionnaires “equally or more acceptable than the usual clinical interview or a written questionnaire” [9].

Other identified barriers to acceptability include cognition, motivation, and perception [48]. For example, one 2003 study of elderly persons with cognitive disabilities found that people did not use 15% of the devices they owned mostly because they did not fit their needs [49]. In order for CAHTs to be adopted, all stakeholders need to be involved in the design and development process. Research suggests that involving users in the design and development of a new system will improve the system’s quality and result in a higher level of user acceptance [50]. To this point, there have been an increasing number of initiatives focused on this approach to engage elderly. The Dutch National Care for the Elderly Programme was one such initiative that put the needs of elderly people at the heart of the program, ensuring their active participation to provide person-centered and integrated care better suited to their needs [51]. The European Commission has also set up a group of measures to improve e-accessibility for older people as part of the e-Inclusion policies [52]. Through user-centered design and educational training programs, accessibility and adoption can be greatly increased.

Considerations and barriers to use from a provider’s perspective include (1) availability of practice workflows and protocols related to patient-generated health data and (2) data storage, accessibility, and ease of use at the point of care [53].

Privacy and Data Security

Acceptability is also largely attributable to fears regarding privacy and data security. The concerns are often related to who has access to sensitive personal health information and to breaches from external malicious counterparts. To address data privacy concerns, there have been updates to global data protection legislation. The General Data Protection Regulation, enforced from 2018, is one such example that strengthens data protection for all individuals within the European Union. The Health Insurance Portability and Accountability Act, enacted in 1996, safeguards the collection, storage, and disclosure of identifiable health data in the United States [54]. Despite these regulatory changes, health care is one of the most data-rich industries globally, with approximately 30% of the world’s electronic data storage occupied by health care information [55]. A person is estimated to accrue more than 1 million gigabytes of health-related data in their lifetime through EHRs, digitized diagnostics, and wearable medical devices [55]. This deluge of health data and the potential value attributable to ransom or sales on the black market makes health data one of the most targeted sources to cyber threats. In 2018, health care data breaches accounted for 24% of all investigated breaches across all industries [56]. In response to this increasing cybercrime, the global cybersecurity market, estimated to be US $100 billion in 2017, is expected to grow to US $173 billion in 2022 at a compound annual growth rate of 11.6% [57]. New technologies are emerging to market to address these security concerns including cryptography (translating data into code, only accessible with an access key), tokenization (sensitive data are substituted with a randomly generated value or token), and distributed ledgers or blockchain (a distributed list of records or blocks that are linked using cryptography). Estonia is an example of an early adopter of this technology utilizing blockchain technology to secure health care data and process transactions [58]. This puts patients in control of their data and allows them to grant or restrict access to different groups including their health care provider, family members, or research teams. With greater control of personal data, there is a greater sense of security.

Although advances in technology are fortifying against data breaches, no system is infallible. The question should be whether the potential risk outweighs the perceived benefit. Logically, people are not willing to share personal information digitally, risking confidentiality, if there is no perceived benefit to them. Historically, this may have been true, as digital health data were largely collected, but underutilized. However, with advances in computing power and artificial intelligence, these volumes of data can be analyzed using predictive models, to provide more accurate and personalized health care. Artificial intelligence has been shown to effectively diagnose and predict multiple conditions by leveraging machine learning, NLP, and

image recognition [59]. It has been shown to predict hospitalization due to heart disease roughly a year in advance with an accuracy of 82% [60]. The pervasive use and continuously improving algorithms employed by artificial intelligence in health care will result in greater realized benefit, largely outweighing the potential risks. The benefits of personalization over data privacy are evident in the consumer industry, where 57% of consumers are willing to share personal data in exchange for personalized offers or discounts [61]. Therefore, how much greater would the benefit of longevity and good health be?

Conclusion

Patient health information is increasingly collected through multiple modalities, including EHRs, wearables, and connected devices. CAHT could provide an additional channel to collect highly relevant, comprehensive, and accurate patient information while reducing the burden on clinicians and face-to-face consultation time. Barriers to implementation and use in practice, such as accessibility, affordability, accuracy, and acceptability, may be addressed by advances in technology, computing power, and ubiquity of personal devices. Thus, perhaps there is no better time than now to adopt CAHT in standard care.

Acknowledgments

There was no funding support for this article.

Authors' Contributions

PS conceived of the idea, researched and drafted the paper, and made subsequent revisions. MK, JA, and NC supervised and contributed to editing and review of the article.

Conflicts of Interest

None declared.

References

1. Ohm F, Vogel D, Sehner S, Wijnen-Meijer M, Harendza S. Details acquired from medical history and patients' experience of empathy--two sides of the same coin. BMC Med Educ 2013 May 09;13(1):67 [FREE Full text] [doi: 10.1186/1472-6920-13-67] [Medline: 23659369]
2. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties. Ann Intern Med 2016 Sep 06;165(11):753. [doi: 10.7326/m16-0961]
3. Elmore N, Burt J, Abel G, Maratos FA, Montague J, Campbell J, et al. Investigating the relationship between consultation length and patient experience: a cross-sectional study in primary care. Br J Gen Pract 2016 Oct 24;66(653):e896-e903. [doi: 10.3399/bjgp16x687733]
4. Howie J, Porter AM, Heaney DJ, Hopton JL. Long to short consultation ratio: a proxy measure of quality of care for general practice. Br J Gen Pract 1991 Feb;41(343):48-54 [FREE Full text] [Medline: 2031735]
5. Dugdale D, Epstein R, Pantilat SZ. Time and the patient-physician relationship. J Gen Intern Med 1999 Jan;14 Suppl 1:S34-S40 [FREE Full text] [doi: 10.1046/j.1525-1497.1999.00263.x] [Medline: 9933493]
6. Irving G, Neves AL, Dambha-Miller H, Oishi A, Tagashira H, Verho A, et al. International variations in primary care physician consultation time: a systematic review of 67 countries. BMJ Open 2017 Nov 08;7(10):e017902 [FREE Full text] [doi: 10.1136/bmjopen-2017-017902] [Medline: 29118053]
7. Hearts: technical package for cardiovascular disease management in primary health care. World Health Organization. 2016. URL: https://www.who.int/cardiovascular_diseases/hearts/Hearts_package.pdf [accessed 2021-02-21]
8. Williams CA. Usability of a Computer-assisted Interview System for the Unaided Self-entry of Patient Data in an Urban Rheumatology Clinic. Journal of the American Medical Informatics Association 2004 Apr 02;11(4):249-259. [doi: 10.1197/jamia.m1527]
9. Pappas Y, Anandan C, Liu J, Car J, Sheikh A, Majeed A. Computer-assisted history-taking systems (CAHTS) in health care: benefits, risks and potential for further development. Inform Prim Care 2011 Jun 01;19(3):155-160 [FREE Full text] [doi: 10.14236/jhi.v19i3.808] [Medline: 22688224]
10. Arora S, Goldberg A, Menchine M. Patient impression and satisfaction of a self-administered, automated medical history-taking device in the Emergency Department. West J Emerg Med 2014 Feb;15(1):35-40 [FREE Full text] [doi: 10.5811/westjem.2013.2.11498] [Medline: 24695871]
11. Ammenwerth E, Schnell-Inderst P, Hoerbst A. Patient empowerment by electronic health records: first results of a systematic review on the benefit of patient portals. Stud Health Technol Inform 2011;165:63-67. [Medline: 21685587]
12. Wolford G, Rosenberg SD, Rosenberg HJ, Swartz MS, Butterfield MI, Swanson JW, et al. A clinical trial comparing interviewer and computer-assisted assessment among clients with severe mental illness. Psychiatr Serv 2008 Jul;59(7):769-775. [doi: 10.1176/ps.2008.59.7.769] [Medline: 18586994]

13. Tompkins BM, Tompkins WJ, Loder E, Noonan AF. A Computer-Assisted Preanesthesia Interview. Anesthesia & Analgesia 1980;59(1):3-10. [doi: 10.1213/00000539-198001000-00002]
14. Bachman J. Improving care with an automated patient history. Fam Pract Manag 2007;14(7):39-43 [FREE Full text] [Medline: 17696057]
15. Bachman JW. The patient-computer interview: a neglected tool that can aid the clinician. Mayo Clin Proc 2003 Jan;78(1):67-78. [doi: 10.4065/78.1.67] [Medline: 12528879]
16. Pappas Y, Wei I, Car J, Majeed A, Sheikh A. Computer-assisted versus oral-and-written family history taking for identifying people with elevated risk of type 2 diabetes mellitus. Cochrane Database Syst Rev 2011 Dec 07(12):CD008489. [doi: 10.1002/14651858.CD008489.pub2] [Medline: 22161431]
17. Zimny NJ, Tandy CJ. Development of a computer-assisted method for the collection, organization, and use of patient health history information in physical therapy. J Orthop Sports Phys Ther 1993 Feb;17(2):84-89. [doi: 10.2519/jospt.1993.17.2.84] [Medline: 8467338]
18. Ancker J, Mauer E, Kalish R, Vest J, Gossey J. Early Adopters of Patient-Generated Health Data Upload in an Electronic Patient Portal. Appl Clin Inform 2019 Mar 10;10(2):254-260 [FREE Full text] [doi: 10.1055/s-0039-1683987] [Medline: 30970383]
19. Radicati S. Mobile Statistics Report, 2014-2018. The Radicati Group, Inc. 2014. URL: https://www.radicati.com/wp/wp-content/uploads/2014/01/Mobile-Statistics-Report-2014-2018-Executive-Summary.pdf [accessed 2021-02-21]
20. Wiggers K. Pew: Smartphone penetration ranges from 24% in India to 95% in South Korea. VentureBeat. 2019. URL: https://venturebeat.com/2019/02/05/pew-south-korea-has-the-worlds-highest-smartphone-ownership-rate/ [accessed 2021-02-21]
21. Smartphone users worldwide 2016-2021. Statista. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2019-02-12]
22. Global mobile consumer trends: Second edition. Deloitte. URL: https://www2.deloitte.com/us/en/pages/technology-media-and-telecommunications/articles/global-mobile-consumer-trends.html [accessed 2021-02-21]
23. Manduchi R, Coughlan J. Portable and Mobile Systems in Assistive Technology. In: Miesenberger K, Klaus J, Zagler W, Karshmer A, editors. Computers Helping People with Special Needs. ICCHP 2008. Lecture Notes in Computer Science, vol 5105. Berlin, Germany: Springer Publishing Company; 2008.
24. Sarji DK. HandTalk: Assistive Technology for the Deaf. Computer 2008 Jul;41(7):84-86. [doi: 10.1109/mc.2008.226]
25. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable Health Devices in Health Care: Narrative Systematic Review. JMIR Mhealth Uhealth 2020 Nov 09;8(11):e18907 [FREE Full text] [doi: 10.2196/18907] [Medline: 33164904]
26. Reisman M. EHRs: The Challenge of Making Electronic Data Usable and Interoperable. P T 2017 Sep;42(9):572-575 [FREE Full text] [Medline: 28890644]
27. Gamache R, Kharrazi H, Weiner J. Public and Population Health Informatics: The Bridging of Big Data to Benefit Communities. Yearb Med Inform 2018 Aug 29;27(1):199-206 [FREE Full text] [doi: 10.1055/s-0038-1667081] [Medline: 30157524]
28. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752 [FREE Full text] [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
29. Wang SJ, Middleton B, Prosser LA, Bardon CG, Spurr CD, Carchidi PJ, et al. A cost-benefit analysis of electronic medical records in primary care. The American Journal of Medicine 2003 Apr;114(5):397-403. [doi: 10.1016/s0002-9343(03)00057-3]
30. How much is this going to cost me? The Office of the National Coordinator for Health Information Technology (ONC). 2014 Nov 12. URL: https://www.healthit.gov/faq/how-much-going-cost-me [accessed 2021-02-21]
31. Electronic Medical Records/Electronic Health Records (EMRs/EHRs). Centers for Disease Control and Prevention. 2020 Mar 3. URL: https://www.cdc.gov/nchs/fastats/electronic-medical-records.htm [accessed 2021-02-21]
32. Schoen C, Osborn R, Squires D, Doty M, Rasmussen P, Pierson R, et al. A survey of primary care doctors in ten countries shows progress in use of health information technology, less in other areas. Health Aff (Millwood) 2012 Dec;31(12):2805-2816. [doi: 10.1377/hlthaff.2012.0884] [Medline: 23154997]
33. The top 100 EHR companies (Part 1 of 4). Medical Economics. 2013 Oct 25. URL: https://www.medicaleconomics.com/medical-economics/news/top-100-ehr-companies-part-1-4 [accessed 2021-02-21]
34. Mobile devices are too expensive for billions of people — and it's keeping them offline. Alliance for Affordable Internet. 2020 Aug 05. URL: https://a4ai.org/mobile-devices-are-too-expensive-for-billions-of-people-and-its-keeping-them-offline/ [accessed 2021-02-21]
35. The 17 Goals. United Nations Department of Economic and Social Affairs. 2015. URL: https://sdgs.un.org/goals [accessed 2021-02-21]
36. Zakim D, Braun N, Fritz P, Alscher MD. Underutilization of information and knowledge in everyday medical practice: evaluation of a computer-based solution. BMC Med Inform Decis Mak 2008 Nov 05;8(1):50 [FREE Full text] [doi: 10.1186/1472-6947-8-50] [Medline: 18983684]

37. Zhou L, Blackley SV, Kowalski L, Doan R, Acker WW, Landman AB, et al. Analysis of Errors in Dictated Clinical Documents Assisted by Speech Recognition Software and Professional Transcriptionists. JAMA Netw Open 2018 Jul 06;1(3):e180530 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.0530] [Medline: 30370424]
38. Chung A, Basch EM. Incorporating the patient's voice into electronic health records through patient-reported outcomes as the "review of systems". J Am Med Inform Assoc 2015 Jul;22(4):914-916 [FREE Full text] [doi: 10.1093/jamia/ocu007] [Medline: 25614143]
39. Islam MM, Topp L, Conigrave KM, van Beek I, Maher L, White A, et al. The reliability of sensitive information provided by injecting drug users in a clinical setting: clinician-administered versus audio computer-assisted self-interviewing (ACASI). AIDS Care 2012 Mar 28;24(12):1496-1503. [doi: 10.1080/09540121.2012.663886] [Medline: 22452446]
40. Mears M, Coonrod DV, Bay RC, Mills TE, Watkins MC. Routine history as compared to audio computer-assisted self-interview for prenatal care history taking. J Reprod Med 2005 Sep;50(9):701-706. [Medline: 16363759]
41. Schouten BC, Meeuwesen L. Cultural differences in medical communication: a review of the literature. Patient Educ Couns 2006 Dec;64(1-3):21-34. [doi: 10.1016/j.pec.2005.11.014] [Medline: 16427760]
42. Tao J, Tan T. Affective Computing: A Review. In: Tao J, Tan T, Picard RW, editors. Affective Computing and Intelligent Interaction. ACII 2005. Lecture Notes in Computer Science, vol 3784. Berlin, Germany: Springer Publishing Company; 2005.
43. Borza D, Danescu R, Itu R, Darabant A. High-Speed Video System for Micro-Expression Detection and Recognition. Sensors (Basel) 2017 Dec 14;17(12):2913 [FREE Full text] [doi: 10.3390/s17122913] [Medline: 29240700]
44. Oh Y, See J, Le Ngo AC, Phan RC, Baskaran VM. A Survey of Automatic Facial Micro-Expression Analysis: Databases, Methods, and Challenges. Front Psychol 2018 Jul 10;9:1128 [FREE Full text] [doi: 10.3389/fpsyg.2018.01128] [Medline: 30042706]
45. Meservy T, Jensen M, Kruse J, Twitchell D, Tsechpenakis G, Burgoon J, et al. Deception Detection through Automatic, Unobtrusive Analysis of Nonverbal Behavior. IEEE Intell. Syst 2005 Sep;20(5):36-43. [doi: 10.1109/mis.2005.85]
46. Gohil S, Vuik S, Darzi A. Sentiment Analysis of Health Care Tweets: Review of the Methods Used. JMIR Public Health Surveill 2018 Apr 23;4(2):e43 [FREE Full text] [doi: 10.2196/publichealth.5789] [Medline: 29685871]
47. Wei I. Computer-assisted versus oral-and-written dietary history taking for diabetes mellitus. Cochrane Database Syst Rev(12): p. Cd00 2011:8488. [doi: 10.1002/14651858.cd008488.pub2]
48. Wildenbos G, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
49. Mann WC, Goodall S, Justiss MD, Tomita M. Dissatisfaction and nonuse of assistive devices among frail elders. Assist Technol 2002 Dec 31;14(2):130-139. [doi: 10.1080/10400435.2002.10132062] [Medline: 14651251]
50. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
51. Jong BM, Wynia K, Geluk-Bleumink A. Ageing Better in the Netherlands. London, England: IntechOpen Limited; 2018.
52. 52007SC1469 - EN. EUR-Lex. 2007. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52007SC1469 [accessed 2021-02-21]
53. Cohen DJ, Keller SR, Hayes GR, Dorr DA, Ash JS, Sittig DF. Integrating Patient-Generated Health Data Into Clinical Care Settings or Clinical Decision-Making: Lessons Learned From Project HealthDesign. JMIR Hum Factors 2016 Oct 19;3(2):e26 [FREE Full text] [doi: 10.2196/humanfactors.5919] [Medline: 27760726]
54. Gostin LO, Halabi SF, Wilson K. Health Data and Privacy in the Digital Era. JAMA 2018 Jul 17;320(3):233-234. [doi: 10.1001/jama.2018.8374] [Medline: 29926092]
55. Sensmeier J. Harnessing the power of artificial intelligence. Nurs Manage 2017 Nov;48(11):14-19. [doi: 10.1097/01.NUMA.0000526062.69220.41] [Medline: 29023284]
56. Jalali MS, Razak S, Gordon W, Perakslis E, Madnick S. Health Care and Cybersecurity: Bibliometric Analysis of the Literature. J Med Internet Res 2019 Feb 15;21(2):e12644 [FREE Full text] [doi: 10.2196/12644] [Medline: 30767908]
57. Annex A-3: Cyber Security. Infocomm Media Development Authority. URL: https://www.imda.gov.sg/-/media/Imda/Files/Industry-Development/Infrastructure/Technology/Technology-Roadmap/WG3-Cyber-Security-Executive-Summary.pdf [accessed 2021-02-21]
58. Healthcare. e-estonia. URL: https://e-estonia.com/solutions/healthcare/e-health-record/ [accessed 2021-02-21]
59. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan 7;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
60. Dai W, Brisimi TS, Adams WG, Mela T, Saligrama V, Paschalidis IC. Prediction of hospitalization due to heart diseases by supervised learning methods. Int J Med Inform 2015 Mar;84(3):189-197 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.10.002] [Medline: 25497295]
61. Digital Advertising 2020. Salesforce. 2018. URL: https://www.salesforce.com/form/marketingcloud/digital-ads-2020-research/ [accessed 2021-02-21]

Abbreviations

CAHT: computer-assisted history taking
CAHTS: computer-assisted history taking system
EHR: electronic health record
NLP: natural language processing

Edited by G Eysenbach; submitted 13.04.20; peer-reviewed by H Spallek, E Metting, C Rey-Reñones, R Chan; comments to author 26.10.20; revised version received 23.12.20; accepted 24.01.21; published 24.02.21.

Please cite as:
Spinazze P, Aardoom J, Chavannes N, Kasteleyn M
The Computer Will See You Now: Overcoming Barriers to Adoption of Computer-Assisted History Taking (CAHT) in Primary Care
J Med Internet Res 2021;23(2):e19306
URL: https://www.jmir.org/2021/2/e19306
doi:10.2196/19306
PMID:33625360

Original Paper

Negative and Positive Affect Regulation in a Transdiagnostic Internet-Based Protocol for Emotional Disorders: Randomized Controlled Trial

Amanda Díaz-García[1], PhD; Alberto González-Robles[1], PhD; Azucena García-Palacios[2,3], PhD; Javier Fernández-Álvarez[4], MSc; Diana Castilla[3,5], PhD; Juana María Bretón[2], PhD; Rosa María Baños[3,5], PhD; Soledad Quero[2], PhD; Cristina Botella[2,3], PhD

[1]Department of Psychology and Sociology, Universidad de Zaragoza, Teruel, Spain

[2]Universitat Jaume I, Castellón de la Plana, Spain

[3]CIBER Fisiopatología Obesidad y Nutrición (CIBERObn), Instituto Carlos III, Madrid, Spain

[4]Department of Psychology, Università Cattolica del Sacro Cuore, Milan, Italy

[5]Department of Personality, Evaluation and Psychological Treatments, Universidad de Valencia, Valencia, Spain

Corresponding Author:
Amanda Díaz-García, PhD
Department of Psychology and Sociology
Universidad de Zaragoza
Calle Cdad. Escolar, S/N, 44003 Teruel
Teruel, 44003
Spain
Phone: 34 878618154
Email: amandadiaz@unizar.es

Abstract

Background: Emotional disorders (EDs) are among the most prevalent mental disorders. Existing evidence-based psychological treatments are not sufficient to reduce the disease burden of mental disorders. It is therefore essential to implement innovative solutions to achieve a successful dissemination of psychological treatment protocols, and in this regard, the use of information and communication technologies such as the internet can be very useful. Furthermore, the literature suggests that not everyone with an ED receives the appropriate treatment. This situation has led to the development of new intervention proposals based on the transdiagnostic perspective, which attempts to address the underlying processes common to EDs. Most of these transdiagnostic interventions focus primarily on downregulating negative affectivity (NA), and less attention has been paid to strengths and the upregulation of positive affectivity, despite its importance for well-being and mental health.

Objective: This study aims to evaluate the efficacy of a transdiagnostic internet-based treatment for EDs in a community sample.

Methods: A 3-armed randomized controlled trial was conducted. A total of 216 participants were randomly assigned to a transdiagnostic internet-based protocol (TIBP), a TIBP+ positive affect (PA) component, or a waiting list (WL) control group. The treatment protocol contained core components mainly addressed to downregulate NA (ie, present-focused emotional awareness and acceptance, cognitive flexibility, behavioral and emotional avoidance patterns, and interoceptive and situational exposure) as well as a PA regulation component to promote psychological strengths and enhance well-being. Data on depression, anxiety, quality of life, neuroticism and extraversion, and PA/NA before and after treatment were analyzed. Expectations and opinions of treatment were also analyzed.

Results: Within-group comparisons indicated significant pre-post reductions in the two experimental conditions. In the TIBP+PA condition, the effect sizes were large for all primary outcomes (d=1.42, Beck Depression Inventory [BDI-II]; d=0.91, Beck Anxiety Inventory [BAI]; d=1.27, Positive and Negative Affect Schedule-Positive [PANAS-P]; d=1.26, Positive and Negative Affect Schedule-Negative [PANAS-N]), whereas the TIBP condition yielded large effect sizes for BDI-II (d=1.19) and PANAS-N (d=1.28) and medium effect sizes for BAI (d=0.63) and PANAS-P (d=0.69). Between-group comparisons revealed that participants who received one of the two active treatments scored better at posttreatment than WL participants. Although there were no statistically significant differences between the two intervention groups on the PA measure, effect sizes were consistently larger in the TIBP+PA condition than in the standard transdiagnostic protocol.

Conclusions: Overall, the findings indicate that EDs can be effectively treated with a transdiagnostic intervention via the internet, as significant improvements in depression, anxiety, and quality of life measures were observed. Regarding PA measures, promising effects were found, but more research is needed to study the role of PA as a therapeutic component.

Trial Registration: ClinicalTrials.gov NCT02578758; https://clinicaltrials.gov/ct2/show/NCT02578758

International Registered Report Identifier (IRRID): RR2-10.1186/s12888-017-1297-z

(J Med Internet Res 2021;23(2):e21335) doi:10.2196/21335

KEYWORDS

transdiagnostic; positive affectivity; negative affectivity; emotion regulation; emotional disorders; internet

Introduction

Transdiagnostic Treatments for the Common Psychopathological Processes Underlying Emotional Disorders

Emotional disorders (EDs) are defined as anxiety and unipolar mood disorders. These disorders have been grouped based on their common biological and psychological vulnerabilities [1]. The estimated lifetime prevalence rates for EDs are high (28.8% for anxiety disorders and 20.8% for mood disorders). In addition, the co-occurrence of multiple EDs has also been found to be elevated, with studies showing that more than 40% of people with one diagnosis also met the diagnostic criteria for a second disorder over a 12-month period [2].

In recent years, research has demonstrated that evidence-based psychological treatments (EBTs) are effective in the treatment of EDs [3]. However, there has been little success in decreasing the prevalence and incidence of mental illness, and only a small proportion of people in need actually receive adequate psychological treatment [4]. In addition, disseminating EBTs has become a real challenge because of their cost, the duration of the treatments, and the lack of well-qualified professionals [5], which can explain why EBTs are underutilized in clinical practice settings [6].

Recently, transdiagnostic approaches have emerged that address the common characteristics found in cognitive, behavioral, emotional, and other dysregulation areas underlying different EDs, that is, the biological and psychological vulnerabilities shared by different mental disorders [7,8]. With regard to transdiagnostic processes, maladaptive emotion regulation strategies have been suggested as potential explanatory factors underlying the comorbidity across EDs [9].

In response to transdiagnostic approaches, several transdiagnostic treatments have been developed to provide patients with a set of skills geared specifically toward common vulnerabilities [10]. One example of these treatments is the Unified Protocol (UP) [7], which was designed to be applicable across different EDs and represented a significant shift toward transdiagnostic psychological treatments for EDs [11,12]. The UP has been tested and results indicate that it is effective in reducing negative affect (NA) [13], with improvements maintained at the 18-month follow-up [14]. Furthermore, the effect of the UP has been shown on the two temperament dimensions of neuroticism (N)/behavioral inhibition (BI) and extraversion (E)/behavioral activation (BA) [15].

Mounting evidence demonstrates the efficacy of transdiagnostic treatments in patients with EDs compared with control groups [16-19], showing that transdiagnostic treatments are just as effective as disorder-specific cognitive behavioral therapy (CBT) [12,20]. The data suggest that a transdiagnostic treatment for EDs might be more widely effective across a diverse range of mental disorders, addressing different disorders with a single protocol [21]. More specifically, a recent meta-analysis showed that the UP is more effective compared with different control groups, such as treatment as usual, waitlist, and medication control groups, in treating anxiety and depressive symptoms [22].

The Role of Positive Affect in EDs

Regarding the temperamental vulnerabilities, some authors have identified two essential dimensions of temperament in the etiology and course of EDs: N/NA and E/positive affect (PA) [8]. Hence, neuroticism has been identified as a core factor involved in the development of EDs [23]. In addition, N/NA and E/PA have been closely related to Gray's (1987) constructs of BI and BA, respectively [24-26], and these terms are often used interchangeably as the most stable measures of temperament [8,15,26,27]. Thus, people with EDs have higher levels of N/NA/BI [8], and they experience negative emotions more intensely and frequently [28] than people who do not have any ED. In contrast, the dimension of positive emotionality, E/PA/BA, has also been observed in many disorders, suggesting that people with an ED show low levels of E/BA [29], which can predict the onset of depression [30] and increase the severity of the problem [31]. Despite the importance of PA in health and well-being, there is limited research on its promotion; therefore, more research is needed in this area. Furthermore, notwithstanding the recent upsurge in transdiagnostic treatments for EDs, most of these protocols have focused on reducing NA. They have addressed core psychopathological deficits in the way patients experience and respond to negative emotions [32]. However, less attention has been paid to positive emotions or promoting PA [33]. In addition to being involved in the symptomatology of EDs, positive emotionality is considered a core element of mental health, showing beneficial, generalized effects on health and functioning [34-37]. Thus, the relationship between emotion regulation (eg, cognitive reappraisal) and well-being has also been demonstrated [38]. On the basis of the literature that highlights the potential importance of positive emotionality as a treatment component [39-43], it is necessary to develop and test treatment components focused on upregulating PA.

Internet-Based Treatments

There are many models for delivering interventions in novel ways that can be scaled up to reach large numbers of people in need [4]. In this regard, information and communication technologies (ICTs) play an important role and can facilitate the availability of EBTs [44]. Specifically, the internet is used for the assessment and treatment of clinical conditions, and it has been established as a useful and effective tool for delivering psychological treatments to treat several psychological disorders [45], particularly depression and anxiety disorders [46]. Moreover, some meta-analyses have revealed that these interventions are as efficacious as face-to-face traditional treatments [47,48].

This Study

The purpose of this study is to test the efficacy of a web-based psychological treatment protocol for individuals from a community sample with one or more diagnoses of EDs: major depressive disorder (MDD), dysthymic disorder (DD), obsessive-compulsive disorder, and four anxiety disorders: panic disorder (PD), agoraphobia (AG), generalized anxiety disorder (GAD), social anxiety disorder (SAD), anxiety disorder not otherwise specified, and (unipolar) mood disorder not otherwise specified [49]. Rather than focusing solely on NA, the treatment protocol includes 2 types of components: one based on classical perspectives for downregulating NA and the other aimed at upregulating PA. The protocol can be applied either in its traditional format (transdiagnostic internet-based protocol, TIBP) or by including both of these components (TIBP+PA).

Some studies have tested the efficacy of transdiagnostic interventions in improving PA measures. However, these studies do not include a specific component to address PA regulation [13], or they are uncontrolled trials [50-52]. Only one study that evaluated the efficacy of a new transdiagnostic treatment focuses on PA, but it is a pilot study rather than a randomized controlled trial (RCT) [53]. To the best of our knowledge, no published RCT has tested the efficacy of a transdiagnostic internet-based treatment for EDs with a specific component to address PA regulation. Therefore, the aim of this study is to investigate the effectiveness of this transdiagnostic protocol for EDs, with and without the specific component to upregulate PA, versus a wait-list control group. A secondary aim is to test the differential effect of the specific treatment component designed to upregulate PA. Finally, we study patients' acceptance of the program developed to apply the treatment protocol over the internet with minimal support by the clinician. We hypothesized that (1) both self-applied protocol modalities (TIBP and TIBP+PA) would be more effective than the wait-list control condition in the treatment of EDs; (2) both interventions would result in significant improvements in depressive and anxious symptomatology at posttreatment; (3) the TIBP+PA would significantly outperform the TIBP group on PA measures; and (4) both protocols are well accepted, with no statistical differences between conditions.

Methods

Study Design

This study was a three-armed superiority RCT in which participants were randomly allocated to 1 of 3 conditions: (1) TIBP, (2) TIBP+PA, and (3) waiting list (WL) control condition. For ethical reasons, participants in the control condition were offered the possibility of receiving the treatment protocol after spending time on the WL (16 weeks), thus leaving no control group for the follow-up measurements. Block randomization was performed to ensure that all primary diagnoses were equally represented across conditions. The trial was registered at ClinicalTrial.gov as NCT02578758 on October 16, 2015. The study was approved by the Ethics Committee of Universitat Jaume I (Castellón, Spain; May 5, 2016) and was conducted in compliance with the study protocol, following the Consolidated Standards of Reporting Trials (CONSORT) statement [54], the CONSORT-eHealth guidelines [55], and the Standard Protocol Items: Recommendations for Interventional Trials guidelines [56,57]. Details of the study protocol have been reported elsewhere [58]. Different effect sizes found in the literature based on the transdiagnostic perspective of EDs were considered to estimate the study power in this study. These calculations were performed with the software program G*Power 3.1 [59] and published in the study protocol [58]. This study reports on pre- to posttreatment data.

Study Population, Recruitment, and Eligibility Criteria

The clinical trial was conducted in a community sample of individuals diagnosed with one or more of the aforementioned disorders. Participants were recruited from adult volunteers interested in participating in the study between June 2015 and July 2018. Potential participants were attended by phone by the clinical team members (who had at least a university master's degree in general health psychology) to explain the study and clarify any doubts. People interested in participating signed the web-based informed consent form and were assessed taking into account all the inclusion criteria. The inclusion criteria were as follows: (1) being at least 18 years old; (2) meeting the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) diagnostic criteria for EDs; (3) having the ability to understand and read Spanish; (4) having access to the internet and an email address; and (5) providing web-based informed consent. The exclusion criteria were as follows: (1) having schizophrenia, bipolar disorder, or alcohol and/or substance dependence disorder; (2) presence of a high risk of suicide (defined by the Mini-International Neuropsychiatry Interview [60] as greater than or equal to 10 points); (3) presence of medical disease/condition that prevents the participant from carrying out the psychological treatment; and (4) receiving another psychological treatment during the study. Receiving pharmacological treatment was not an exclusion criterion, but any increase and/or change in the medication (in the case of receiving) during the study period implied the participant's exclusion from subsequent analyses. Participants who fulfilled all the study criteria were randomized to one of the three experimental conditions by an independent researcher. This researcher was unaware of the characteristics of the study and had no clinical involvement in the trial or access to the study

data. Participants agreed to participate before determining which treatment they were allocated. All participants were free to withdraw from the treatment at any time. Access and participation in the study did not involve payment in any case.

The Transdiagnostic Interventions

The treatment protocol is based on the transdiagnostic perspective derived from the UP [5,7] and some strategies from Marsha Linehan's protocol [61]. Initially, a manualized protocol was developed and structured in a patient and therapist handbook. Later, the protocol was adapted to a multimedia web platform (videos, vignettes, audios, images, etc) to be completely self-applied via the internet [62] through a PC or a tablet. The ease of use of the program has been strengthened because it presents a linear navigation to optimize the treatment structure and make the treatment easier and more attractive to the participants.

The program consists of an assessment protocol and a treatment protocol that includes core components, mainly designed to downregulate NA (present-focused emotional awareness and acceptance, cognitive flexibility, behavioral and emotional avoidance patterns, and interoceptive and situational exposure) and upregulate PA to promote psychological strengths and enhance well-being [63]. The protocol content is adapted from the UP [7] and some of the strategies for emotion regulation from dialectical behavior therapy [61]. The PA regulation component is based mainly on BA strategies [64], strategies to promote pleasant and significant activities linked to values and life goals, and strategies to enhance personal strengths, positive feelings, positive cognitions, and positive behavior [63,65]. Furthermore, well-being therapy strategies [66,67] and some concepts from Fredickson's Broaden-and-Build Theory [68] are also included in the program. The PA regulation component takes place after the NA regulation component. The protocol also includes traditional therapeutic components of evidence-based treatment for ED (psychoeducation, motivation for change, and relapse prevention). All the treatment components were developed through two self-applied protocol modalities (TIBP and TIBP+PA) with 12 and 16 modules, respectively, with the only difference being the inclusion or absence of the modules that contain the PA-regulation component. A detailed description of modules that contain the PA regulation component is presented in Multimedia Appendix 1 [64,66,68-77]. The modules in each intervention protocol are described briefly elsewhere [58].

The duration of the program could vary among users, and participants in both treatment conditions had equal access to the protocol for a maximum period of 18 weeks. The program sent weekly messages to the patient to remind him/her to continue to work to benefit from the program. A professional platform was used to send these messages [78]. The program also sent automatic emails with reminders to access the modules when participants had not entered the past 15 days. In addition to this ICT support, human support was also provided through weekly phone calls (maximum of 5 min) during the treatment period to resolve any difficulties or doubts, or to remind them of the importance of reviewing the treatment contents.

Outcome Measures

The assessment protocol was included at the beginning and end of the web-based program. A detailed description of the measures and their aims has been published elsewhere [58]. The measures included in this study are described in Table 1.

Table 1. Study measures.

Measure	Aim	Cronbach α	ω[a]	Time of assessment
Diagnostic interview				
MINI[b]	Psychiatric diagnosis	N/A[c]	N/A	BL[d]
Primary outcomes				
BDI-II[e]	Severity of depression	.91	0.91	BL, Post-T[f]
BAI[g]	Severity of anxiety	.92	0.92	BL, Post-T
PANAS[h]	Positive and negative affect	PA[i]=.91; NA[j]=.89	PA=0.91; NA=0.89	BL, Post-T
Secondary outcomes				
Personality measures				
NEO FFI[k]	Neuroticism and extraversion	N[l]=.81; E[m]=.84	N=0.82; E=0.84	BL, Post-T
Quality of life				
EQ-5D[n]	Health-related quality of life	.67	0.70	BL, Post-T
Expectation and opinion				
Expectation of treatment scale	Expectation of treatment	N/A	N/A	BL
Opinion of treatment scale	Opinion of treatment	N/A	N/A	Post-T

[a]ω: coefficient omega in this study.
[b]MINI: Mini-International Neuropsychiatric Interview, Version 5.0.0.
[c]N/A: not applicable.
[d]BL: baseline.
[e]BDI-II: Beck Depression Inventory-II.
[f]Post-T: posttreatment.
[g]BAI: Beck Anxiety Inventory.
[h]PANAS: Positive and Negative Affect Schedule.
[i]PA: positive affect.
[j]NA: negative affect.
[k]NEO FFI: NEO Five Factor Inventory.
[l]N: neuroticism.
[m]E: extraversion.
[n]EQ-5D: EuroQoL-5D Questionnaire.

Statistical Analysis

Group differences in participants' sociodemographic and clinical data at baseline were examined to confirm that they were comparable after randomization. One-way analysis of variance for continuous variables and Fisher exact tests of independence for categorical variables were used. Intention-to-treat (ITT) using mixed models, with full information maximum likelihood estimation and without any ad hoc imputations were conducted to handle missing data due to participant dropout [79]. This approach uses all available data, does not substitute missing values with assumed or estimated values, and does not assume that the last measurement is stable (the last observation carried forward assumption) [80]. Mixed model analyses are appropriate for RCTs with multiple time points and pre-to postonly designs with substantial dropout rates [81]. The pattern of missingness was investigated to determine its likelihood of being random rather than systematic (missing not at random, MNAR). Subsequently, associations between sample characteristics missingness in the outcome variables were examined (*t* tests for continuous variables and Fisher exact tests for categorical variables). A linear mixed model for each outcome measure was implemented using the linear mixed-effects models (MIXED) procedure with one random intercept per subject. An identity covariance structure was specified to model the covariance structure of the random intercept. Significant effects were followed up with pairwise comparisons using the Bonferroni correction. Effect sizes were calculated for within- and between-group comparisons using the standardized observed mean difference proposed by Cohen [82]. To determine the existence of a reliable change in a patient, the reliable change index (RCI; Jacobson and Truax's method) [83] was used. The RCI values for the primary outcomes (Beck Depression Inventory, BDI-II; Beck Anxiety Inventory, BAI; Positive and Negative Affect Schedule-Positive [PANAS-P]; and Positive and Negative Affect Schedule-Negative [PANAS-N]) were calculated for the completer sample (participants who provided data at posttreatment). Fisher exact tests were performed to evaluate group differences in RCI rates for completers. All statistical analyses were conducted using IBM SPSS Statistics

for Windows, version 22, and SAS software, version 9.4, of the SAS System for Windows.

Results

Participant Flow and Attrition

Out of the 573 people who expressed initial interest in the study, as the flow diagram shows (Figure 1), only 402 performed the initial interview. At this stage, 186 participants failed to meet the inclusion criteria. Finally, 216 patients were included in the study, and they were randomly allocated to each experimental condition: TIBP, n=71; TIBP+PA, n=73; WL, n=72. Regarding pretreatment assessments, 71 participants performed it in the TIBP, 73 in the TIBP+PA, and 72 in the WL. A similar number of participants performed the posttreatment assessment from both intervention conditions (TIBP, n=45; TIBP+PA, n=46). No significant differences between the three conditions were found in dropout rates (X^2_2=3.8, *P*=.14). In the TIBP condition, of those who started the program (n=71), 26 participants (26/71, 37%) withdrew from the treatment. In the TIBP+PA condition, a similar pattern was found; of those who started the program (n=73), 27 participants (27/73, 37%) withdrew from the treatment. Finally, in the WL control group, data from 55 participants were obtained after they had spent 16 weeks on the WL (55/72, 76% retention; 17/72, 24% dropout). Overall, of the 216 participants who started the study, 86 participants withdrew from the program. As a result, two patterns of missingness emerged. One of them represented 32% of the sample (70/216, 32.4%) and the other represented a very low percentage (16/216, 7.4%). Missingness was not related to the characteristics listed in Table 2 in any of the three arms of the RCT (all *P*>.05 in both patterns). Therefore, patterns of missingness were not found to be MNAR and the decision to continue the analysis with the available data was made [84,85].

Figure 1. Flowchart of participants. DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision; ED: emotional disorder; TIBP: Transdiagnostic internet-Based Protocol; PA: positive affect; WL: Waiting List; ITT: intention-to-treat.

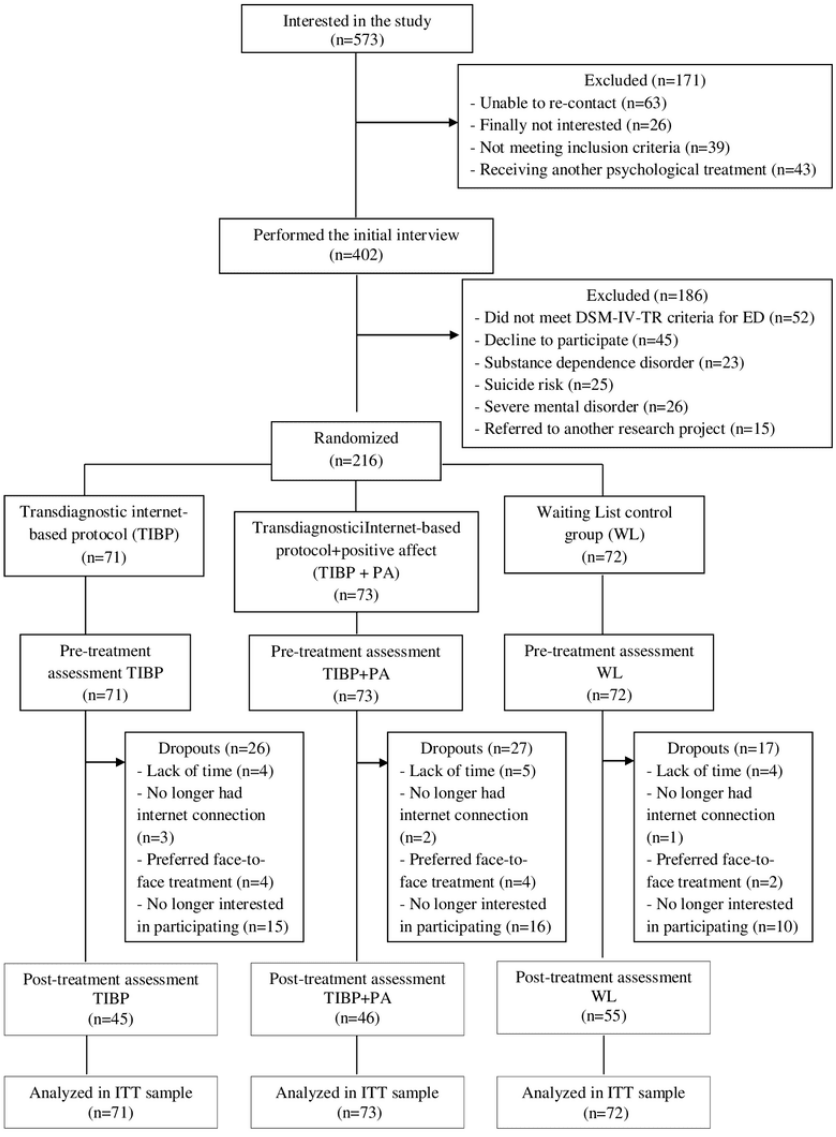


Table 2. Demographic characteristics of participants at pre-assessment (N=216).

Variable	TIBP[a] (n=71)	TIBP+PA[b] (n=73)	WL[c] (n=72)	Total (N=216)	Statistic[d]	*P* value
Age (years), mean (SD); range	35.82 (13.04); 18-72	33.11 (9.74); 19-52	31.82 (10.50); 19-58	33.57 (11.24); 18-72	F_{2213}=2.382	.09
Sex, n (%)					X_2^2=0.2	.88
Female	51 (72)	51 (70)	53 (74)	155 (71.8)		
Male	20 (28)	22 (30)	19 (26)	61 (28.2)		
Marital status, n (%)					N/A[e]	.97
Single	44 (62)	41 (56)	42 (58)	127 (58.8)		
Married or partnered	23 (32)	27 (37)	25 (35)	75 (34.7)		
Divorced or widowed	4 (6)	5 (7)	5 (7)	14 (6.5)		
Education level, n (%)					N/A	.46
Basic studies	1 (1)	5 (7)	3 (4)	9 (4.2)		
Medium studies	15 (21)	12 (16)	17 (24)	44 (20.4)		
Higher studies	55 (78)	56 (77)	52 (72)	163 (75.5)		
Principal diagnosis, n (%)					N/A	.57
MDD[f]	11 (15)	9 (12)	16 (22)	36 (16.7)		
DD[g]	1 (1)	0 (0)	2 (3)	3 (1.4)		
GAD[h]	26 (37)	21 (29)	24 (33)	71 (32.9)		
PD[i]/AG[j]	4 (6)	6 (8)	6 (8)	16 (7.4)		
PD	2 (3)	4 (6)	3 (4)	9 (4.2)		
AG	6 (9)	6 (8)	1 (1)	13 (6.0)		
SAD[k]	15 (21)	23 (32)	16 (22)	54 (25.0)		
OCD[l]	3 (4)	1 (1)	2 (3)	6 (2.8)		
Anxiety NOS[m]	3 (4)	3 (4)	1 (1)	7 (3.2)		
Depression NOS	0 (0)	0 (0)	1 (1)	1 (0.5)		
Number of comorbid disorders, n (%)					N/A	.17
0	32 (45)	19 (26)	34 (47)	81 (37.5)		
1	28 (39)	36 (49)	26 (36)	92 (42.6)		
MDD	18 (64)	22 (61)	15 (57)	57 (62)		
DD	1 (4)	2 (6)	3 (12)	6 (7)		
GAD	2 (7)	7 (19)	2 (8)	11 (12)		
PD/AG	0 (0)	0 (0)	0 (0)	0 (0.0)		
PD	0 (0)	0 (0)	1 (4)	1 (1)		
AG	3 (11)	2 (6)	2 (8)	7 (8)		
SAD	4 (14)	3 (8)	3 (11)	10 (10)		
OCD	0 (0)	0 (0)	0 (0)	0 (0)		
Anxiety NOS	0 (0)	0 (0)	0 (0)	0 (0)		
Depression NOS	0 (0)	0 (0)	0 (0)	0 (0)		
2	7 (10)	12 (16)	7 (10)	27 (12.5)		
MDD and GAD	2 (29)	3 (25)	2 (29)	7 (26)		
MDD and PD/AG	1 (14)	1 (8)	1 (14)	3 (11)		

Variable		TIBP[a] (n=71)	TIBP+PA[b] (n=73)	WL[c] (n=72)	Total (N=216)	Statistic[d]	*P* value
	MDD and PD	1 (14)	1 (8)	2 (29)	4 (15)		
	MDD and AG	0 (0)	2 (17)	0 (0)	3 (11)		
	MDD and OCD	1 (14)	0 (0)	0 (0)	1 (4)		
	DD and SAD	0 (0)	1 (8)	0 (0)	1 (4)		
	GAD and SAD	2 (29)	1 (8)	1 (14)	4 (15)		
	GAD and AG	0 (0)	2 (17)	0 (0)	2 (7)		
	PD/AG and SAD	0 (0)	0 (0)	1 (14)	1 (4)		
	SAD and OCD	0 (0)	1 (8)	0 (0)	1 (4)		
3		4 (6)	6 (8)	5 (7)	16 (7.4)		
	MDD, GAD, and AG	1 (25)	1 (17)	2 (40)	4 (25)		
	MDD, GAD, and SAD	0 (0)	2 (33)	2 (40)	4 (25)		
	MDD, GAD, and OCD	0 (0)	1 (17)	0 (0)	2 (13)		
	MDD, PD, and SAD	0 (0)	2 (33)	0 (0)	2 (13)		
	MDD, PD, and OCD	0 (0)	0 (0)	1 (20)	1 (6)		
	MDD, AG, and SAD	1 (25)	0 (0)	0 (0)	1 (6)		
	MDD, SAD, and OCD	1 (25)	0 (0)	0 (0)	1 (6)		
	GAD, PD/AG, and SAD	1 (25)	0 (0)	0 (0)	1 (6)		

[a]TIBP: transdiagnostic internet-based protocol.

[b]TIBP+PA: transdiagnostic internet-based protocol+positive affect component.

[c]WL: waiting list.

[d]Statistic: Pearson chi-square or Fisher exact test.

[e]N/A: not applicable.

[f]MDD: major depressive disorder.

[g]DD: dysthymic disorder.

[h]GAD: generalized anxiety disorder.

[i]PD: panic disorder.

[j]AG: agoraphobia.

[k]SAD: social anxiety disorder.

[l]OCD: obsessive-compulsive disorder.

[m]NOS: not otherwise specified.

Baseline Data and Participant Characteristics

Details about participants' sociodemographic characteristics for each group at pretreatment are presented in Table 2. The results indicated that there were no significant differences between the experimental groups before treatment for any of these variables, indicating that the randomization was successful. Overall, participants' mean age was 33.57 years (SD 11.24, range 18-72), the majority were females (155/216, 71.8%), and most of them were single (127/216, 58.8%) and had completed or were pursuing higher studies (163/216, 75.5%; eg, undergraduate degree studies, graduate studies or university master's degrees, or postgraduate studies or doctoral degrees).

Principal and comorbid diagnoses are presented in Table 2. Most of the participants had GAD (71/216, 32.9%), followed by SAD (54/216, 25.0%) and MDD (36/216, 16.7%). Regarding the patterns of comorbidity in the sample, 41.7% (90/216) of the participants had at least one comorbid diagnosis, with MDD being the most common comorbid disorder (n=57), followed by GAD (n=11), SAD (n=10), AG (n=7), DD (n=6), and PD (n=1).

Regarding the clinical characteristics of the participants in each experimental condition at pretreatment (Table 3), no statistically significant differences were found between the groups on any of the primary and secondary outcomes.

Table 3. Clinical characteristics of participants at pre-assessment.

Measure	TIBP[a] (n=71), mean (SD)	TIBP+PA[b] (n=73), mean (SD)	WL[c] (n=72), mean (SD)	Total (N=216), mean (SD)	Statistic, *F* (*df*)	*P* value
Primary outcomes						
BDI-II[d]	25.35 (11.33)	29.07 (11.33)	26.31 (12.43)	26.93 (11.76)	1.966 (2213)	.14
BAI[e]	21.14 (12.10)	23.58 (11.50)	21.87 (12.56)	22.23 (12.04)	0.719 (2,197)	.49
PANAS[f]_Positive	20.72 (6.52)	19.32 (6.22)	19.28 (5.67)	19.84 (6.22)	1.277 (2213)	.28
PANAS_Negative	30.68 (7.73)	31.96 (8.79)	28.63 (9.03)	30.43 (8.61)	2.807 (2213)	.06
Secondary outcomes						
Personality measures						
NEO FFI[g]_Neuroticism	31.54 (6.65)	34.11 (8.18)	31.39 (8.20)	32.36 (7.78)	2.853 (2213)	.06
NEO FFI_Extraversion	21.49 (8.77)	19.26 (7.59)	21.79 (8.66)	20.84 (8.39)	1.991 (2213)	.14
Quality of life						
EQ-5D[h]	54.60 (17.76)	54.78 (19.75)	51.91 (17.81)	53.75 (18.44)	0.511 (2197)	.60

[a]TIBP: transdiagnostic internet-based protocol.

[b]TIBP+PA: transdiagnostic internet-based protocol+positive affect component.

[c]WL: waiting list.

[d]BDI-II: Beck Depression Inventory-II.

[e]BAI: Beck Anxiety Inventory.

[f]PANAS: Positive and Negative Affect Schedule.

[g]NEO FFI: NEO Five Factor Inventory.

[h]EQ-5D: EuroQoL-5D questionnaire.

Effectiveness of the Intervention on Primary and Secondary Outcomes at Pre-Post

Primary Outcomes

Table 4 includes descriptive statistics (ie, means and SD) for TIBP+PA, TIBP, and WL at pretreatment and posttreatment; Table 5 includes within-group and between-group effect sizes and CIs for all the primary outcome measures in the three experimental groups, based on the ITT sample.

For the three primary outcomes, within-group comparisons indicated significant pre-post reductions in the two experimental conditions, with large effect sizes for the BDI-II (*d*=1.19) and PANAS-N (*d*=1.28), and moderate effect sizes for the BAI (*d*=0.63) and PANAS-P (*d*=0.69) in the TIBP condition. In the TIBP+PA condition, the effect sizes were large for all primary outcomes (*d*=1.42, BDI-II; *d*=0.91, BAI; *d*=1.27, PANAS-P; *d*=1.26, PANAS-N). Between-group comparisons revealed that participants who received the treatment scored better at posttreatment than the WL group. Greater reductions were found in the BDI-II scores in the TIBP condition than in the WL condition (mean difference −13.61; *P*<.001; *d*=1.18; 95% CI −1.61 to −0.76), as well as between the TIBP+PA condition and WL (mean difference −14.31; *P*<.001; *d*=1.05; 95% CI −1.46 to −0.63), with large effect sizes. No differences were found between the two experimental conditions (mean difference 0.70; *P*=.76; *d*=0.10; 95% CI −0.51 to −0.31).The results for BAI scores were similar to the pattern of findings for the BDI: greater reductions in the TIBP condition (mean difference −8.19; *P*=.001; *d*=0.63; 95% CI −1.07 to −0.20) and TIBP+PA condition (mean difference −9.28; *P*<.001; *d*=0.68; 95% CI −1.10 to −0.26), compared with WL, with medium effect sizes, and no differences between the two experimental conditions (mean difference 1.09; *P*=.65; *d*=0.05; 95% CI −0.39 to 0.49). Finally, patients in the TIBP condition experienced a large increase in PA (PANAS-P) compared with WL (mean difference 5.42; *P*<.001; *d*=0.74; 95% CI 0.33 to 1.15) with moderate effect sizes and greater reductions in NA (PANAS-N; mean difference −8.34; *P*<.001; *d*=0.99; 95% CI −1.41 to −0.57) compared with WL with large effect sizes. Participants in the TIBP+PA condition experienced the same pattern as the participants in the TIBP condition but achieving large effect sizes for both higher PA (mean difference 7.86; *P*<.001; *d*=0.90; 95% CI 0.49 to 1.31) and lower NA (mean difference −8.32; *P*<.001; *d*=0.91; 95% CI −1.32 to −0.50) than participants in the WL condition. No differences were found between the two experimental conditions on PA (mean difference −2.44; *P*=.08; *d*=0.25; 95% CI −0.66 to 0.17) or NA (mean difference −0.02; *P*=.99; *d*=0.01; 95% CI −0.42 to 0.40).

Table 4. Descriptive statistics for TIBP+PA, TIBP, and WL at pretreatment and posttreatment for primary outcomes, personality measures, and quality of life measures.

Measures	TIBP+PA[a] (n=46), mean (SD)		TIBP[b] (n=45), mean (SD)		WL[c] (n=55), mean (SD)	
	Pre-T[d]	Post-T[e]	Pre-T	Post-T	Pre-T	Post-T
Primary outcomes						
BDI-II[f]	29.07 (11.33)	12.78 (10.76)	25.35 (11.33)	11.76 (9.02)	26.31 (12.42)	26.09 (13.99)
BAI[g]	23.58 (11.50)	12.97 (10.33)	21.14 (12.10)	13.46 (10.26)	21.87 (12.56)	21.67 (14.41)
Positive Affect subscale of the Positive and Negative Affect Schedule	19.32 (6.22)	27.28 (9.21)	20.72 (6.52)	25.24 (7.10)	19.28 (5.68)	19.86 (7.27)
Negative Affect subscale of the Positive and Negative Affect Schedule	31.96 (8.79)	20.78 (8.37)	30.68 (7.73)	20.71 (6.81)	28.63 (9.03)	28.76 (8.99)
Secondary outcomes						
Personality measures						
Neuroticism subscale of the NEO FFI[h]	34.11 (8.18)	27.67 (8.37)	31.54 (6.65)	26.64 (7.83)	31.39 (8.20)	31.71 (8.56)
Extraversion subscale of the NEO FFI	19.26 (7.59)	24.22 (7.64)	21.49 (8.77)	24.13 (8.41)	21.79 (8.66)	19.84 (10.02)
Quality of life						
EuroQoL-5D-3L questionnaire	54.78 (19.75)	68.81 (18.37)	54.60 (17.76)	64.05 (18.78)	51.91 (17.81)	52.55 (18.96)

[a]TIBP+PA: transdiagnostic internet-based protocol + positive affect component.

[b]TIBP: transdiagnostic internet-based protocol.

[c]WL: waiting list.

[d]Pre-T: Pre-treatment.

[e]Post-T: Post-treatment.

[f]BDI-II: Beck Depression Inventory-II.

[g]BAI: Beck Anxiety Inventory.

[h]NEO FFI: NEO Five Factor Inventory.

Table 5. Within- and between-group effect sizes and 95% CIs.

Measures	TIBP+PA[a], *d* (95% CI)	TIBP[b], *d* (95% CI)	WL[c], *d* (95% CI)	TIBP+PA versus TIBP, *d* (95% CI)	TIBP+PA versus WL, *d* (95% CI)	TIBP versus WL, *d* (95% CI)
	Pre-post	Pre-post	Pre-post	Post-T[d]	Post-T	Post-T
Primary outcomes						
BDI-II[e]	1.42 (1.09 to 1.76)	1.19 (0.90 to 1.48)	0.02 (−0.13 to 0.16)	−0.10 (−0.51 to 0.31)	−1.05 (−1.46 to −0.63)	−1.18 (−1.61 to −0.76)
BAI[f]	0.91 (0.64 to 1.18)	0.63 (0.38 to 0.87)	0.02 (−0.12 to 0.15)	0.05 (−0.39 to 0.49)	−0.68 (−1.10 to −0.26)	−0.63 (−1.07 to −0.20)
Positive Affect subscale of the Positive and Negative Affect Schedule	−1.27 (−1.58 to −0.95)	−0.69 (−0.92 to −0.45)	−0.01 (−0.22 to 0.20)	−0.25 (−0.66 to 0.17)	0.90 (0.49 to 1.31)	0.74 (0.33 to 1.15)
Negative Affect subscale of the Positive and Negative Affect Schedule	1.26 (0.94 to 1.57)	1.28 (0.96 to 1.60)	−0.04 (−0.22 to 0.14)	−0.01 (−0.42 to 0.40)	−0.91(−1.32 to −0.50)	−0.99 (−1.41 to −0.57)
Secondary outcomes						
Personality measures						
Neuroticism subscale of the NEO FFI[g]	0.78 (0.51 to 1.05)	0.73 (0.50 to 0.96)	0.04 (−0.22 to 0.20)	−0.12 (−0.54 to 0.29)	−0.47 (−0.87 to −0.08)	0.61 (−1.01 to −0.21)
Extraversion subscale of the NEO FFI	0.65 (−0.88 to −0.42)	0.30 (−0.45 to −0.14)	0.22 (0.07 to 0.37)	−0.01 (−0.42 to 0.40)	0.48 (0.08 to 0.88)	0.46 (0.06 to 0.86)
Quality of life						
EuroQoL-5D-3L questionnaire	0.70 (−0.97 to −0.43)	0.53 (−0.80 to −0.25)	0.04 (−0.21 to 0.14)	−0.25 (−0.70 to 0.19)	0.86 (0.44 to 1.29)	0.60 (0.17 to 1.04)

[a]TIBP+PA: transdiagnostic internet-based protocol+positive affect component.

[b]TIBP: transdiagnostic internet-based protocol.

[c]WL: waiting list.

[d]Post-T: posttreatment.

[e]BDI-II: Beck Depression Inventory-II.

[f]BAI: Beck Anxiety Inventory.

[g]NEO FFI: NEO Five Factor Inventory.

Secondary Outcomes

Table 4 includes descriptive statistics (ie, mean and SD) for TIBP+PA, TIBP, and WL at pretreatment and posttreatment, and Table 5 includes within-group and between-group effect sizes and CIs for secondary outcomes related to personality and quality of life measures in the three experimental groups based on the ITT sample.

Regarding personality measures, within-group comparisons indicated a significant pre-to-post reduction in neuroticism in the two experimental conditions, with moderate effect sizes in NEO Five Factor Inventory (NEO FFI)-Neuroticism (d=0.73, TIBP; d=0.78, TIBP+PA), and a significant pre-to-post increase in extraversion in the two experimental conditions, with a small effect size in the TIBP condition (d=0.30) and a moderate effect size in the TIBP+PA condition (d=0.65). In the WL control group, significant changes with small effect sizes were also found on NEO FFI-Extraversion (d=0.22). Between-group comparisons revealed that participants who received the treatment scored better at posttreatment on NEO FFI-Neuroticism in both intervention groups compared with the WL group, with moderate effect sizes (d=0.61, TIBP; d=0.47, TIBP+PA). NEO FFI-Extraversion showed better scores with small effect sizes in both intervention conditions (d=0.46, TIBP; d=0.48, TIBP+PA) than WL. No statistically significant differences were found between the two experimental conditions for the personality measures.

Finally, quality of life measures (ie, EuroQoL-5D Questionnaire; EQ-5D) showed a significant time effect ($F_{1,152.98}$=32.98; P<.001). Both intervention groups experienced significant improvements in quality of life posttreatment, and this improvement was not found in the WL control group. Within-group comparisons indicated moderate effect sizes in the TIBP condition (d=0.53), moderate effect sizes in the TIBP+PA condition (d=0.70), and nonsignificant changes in the WL control group. Between-group comparisons revealed that participants who received the treatment (with or without the specific component to upregulate PA) scored higher on quality of life at posttreatment, compared with the WL, with moderate effect sizes in the TIBP condition (d=0.60) and large effect sizes in the TIBP+PA condition (d=0.86; Table 4 for details). The differences between the two treatment groups were not statistically significant.

Expectations and Satisfaction

Table 6 lists the results for the two interventions groups. Items for expectations and satisfaction were rated from 0 (strongly disagree) to 10 (strongly agree), covering how logical the treatment seemed, to what extent it could satisfy the patient, whether it could be used to treat other psychological problems, and its usefulness for the patient's specific problem. Before the treatment, all the scores were high. The analysis revealed statistically significant differences between the two conditions on expectations about the treatment: before treatment, participants in the TIBP+PA condition considered the treatment more logical ($F_{1,89}$=4.49; *P*=.03), more satisfactory ($F_{1,89}$=6.29; *P*=.01), more recommendable to others ($F_{1,89}$=6.15; *P*=.01), and more useful for other psychological problems ($F_{1,89}$=7.38; *P*=.008) than the TIBP participants did. In addition, at posttreatment, participants' satisfaction scores were also high. The analysis revealed statistically significant differences between the two conditions on satisfaction: after treatment, participants in the TIBP+PA condition considered the treatment more satisfactory ($F_{1,89}$=4.10; *P*=.04), recommendable to others ($F_{1,89}$=6.79; *P*=.01), more useful for other psychological problems ($F_{1,89}$=5.13; *P*=.02), and more useful for the patient ($F_{1,89}$=5.91; *P*=.01) than participants in the TIBP condition did.

Table 6. Means and SDs for expectations and satisfaction.

Statements	Expectations		Satisfaction	
	n	Mean (SD)	n	Mean (SD)
How logical do you think this treatment is?				
Total sample	91	8.21 (1.55)	91	8.21 (1.67)
TIBP[a]	45	7.87 (1.71)	45	8.07 (1.68)
TIBP+PA[b]	46	8.54 (1.31)	46	8.35 (1.66)
How satisfied are you with the treatment received?				
Total sample	91	8.21 (1.75)	91	7.63 (2.02)
TIBP	45	7.76 (1.86)	45	7.20 (2.04)
TIBP+PA	46	8.65 (1.54)	46	8.04 (1.93)
To what extent do you feel confident about recommending this treatment to a friend who has the same problems?				
Total sample	91	8.36 (1.77)	91	8.01 (2.12)
TIBP	45	7.91 (1.95)	45	7.44 (2.32)
TIBP+PA	46	8.80 (1.45)	46	8.57 (1.75)
Do you think this treatment could be useful for treating other psychological disorders?				
Total sample	91	8.03 (1.72)	91	7.88 (1.76)
TIBP	45	7.56 (1.79)	45	7.47 (1.89)
TIBP+PA	46	8.50 (1.52)	46	8.28 (1.53)
To what extent do you think the treatment was helpful to you?				
Total sample	91	7.44 (1.93)	91	7.30 (2.07)
TIBP	45	7.13 (1.87)	45	6.78 (2.00)
TIBP+PA	46	7.74 (1.97)	46	7.80 (2.03)

[a]TIBP: transdiagnostic internet-based protocol.

[b]TIBP+PA: transdiagnostic internet-based protocol+positive affect component.

Reliable and Clinically Significant Change

On the basis of the two criteria proposed by Jacobson and Truax to estimate clinically meaningful improvement, patients were classified as recovered, improved, stable, or deteriorated at posttreatment. The RCI has been expressed graphically to facilitate a more intuitive interpretation (Figures 2-5). At posttreatment, statistically significant differences were found between the three conditions in these percentages for all primary outcomes: BDI-II (*P*<.001), BAI (*P*<.001), PANAS-P (*P*=.001), and PANAS-N (*P*<.001). Overall, participants who received the transdiagnostic internet-based interventions showed a higher percentage of recovery compared with WL, which obtained high percentages of reliable deterioration on all the primary outcomes. No statistically significant differences were found between TIBP+PA and TIBP (BDI-II, *P*=.59; BAI, *P*=.67; PANAS-P, *P*=.59; and PANAS-N, *P*=.59).

Figure 2. Percentages of participants recovered, improved, stable, and deteriorated on depression scores (Beck Depression Inventory-II) in transdiagnostic internet-based protocol+positive affect component (inner circle), transdiagnostic internet-based protocol (middle circle), and waiting list (outer circle).

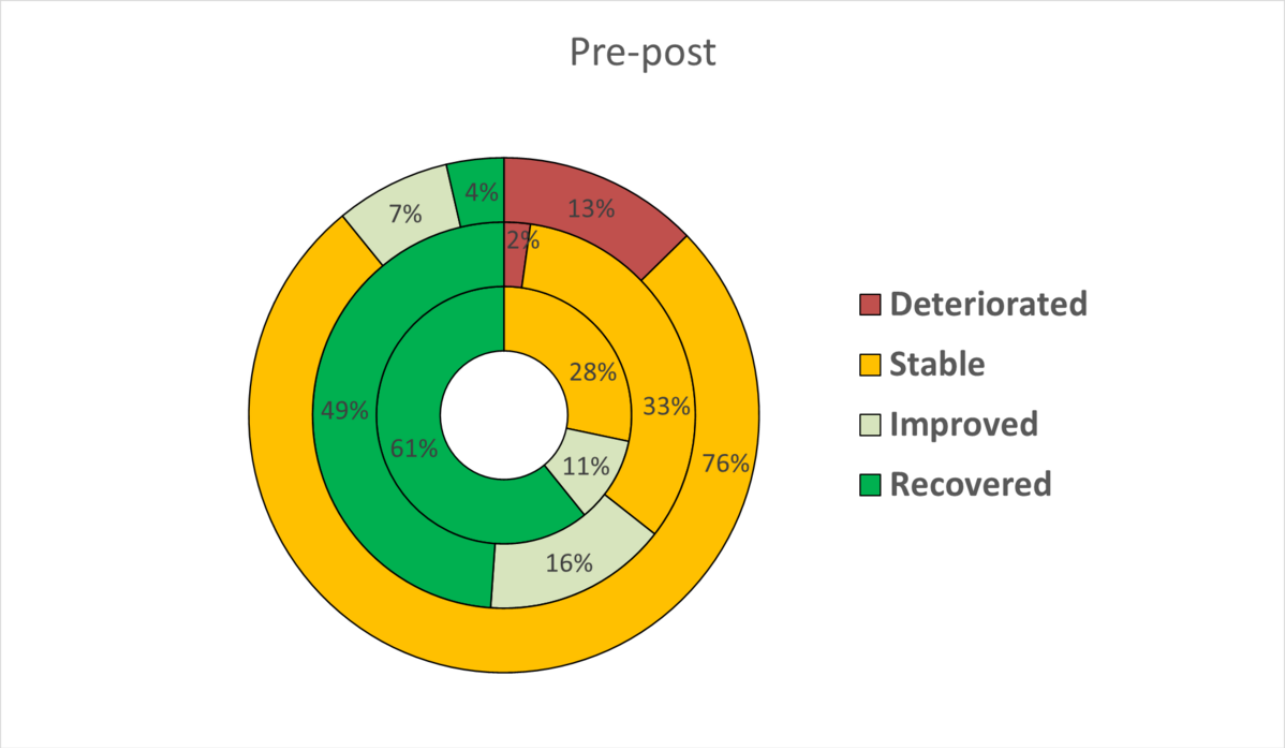


Figure 3. Percentages of participants recovered, improved, stable, and deteriorated on depression scores (Beck Anxiety Inventory) in transdiagnostic internet-based protocol+positive affect component (inner circle), transdiagnostic internet-based protocol (middle circle), and waiting list (outer circle).

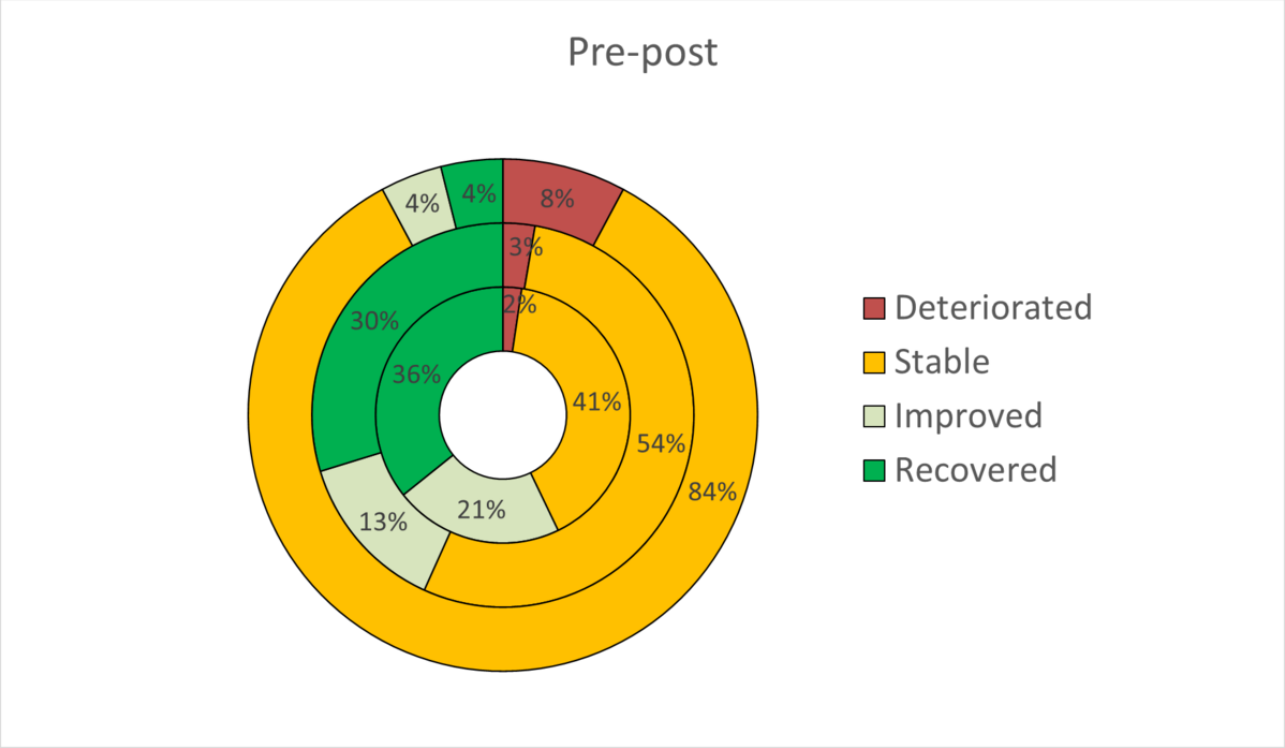


Figure 4. Percentages of participants recovered, improved, stable, and deteriorated on depression scores (Positive and Negative Affect Schedule_Positive subscale) in transdiagnostic internet-based protocol+positive affect component (inner circle), transdiagnostic internet-based protocol (middle circle), and waiting list (outer circle).

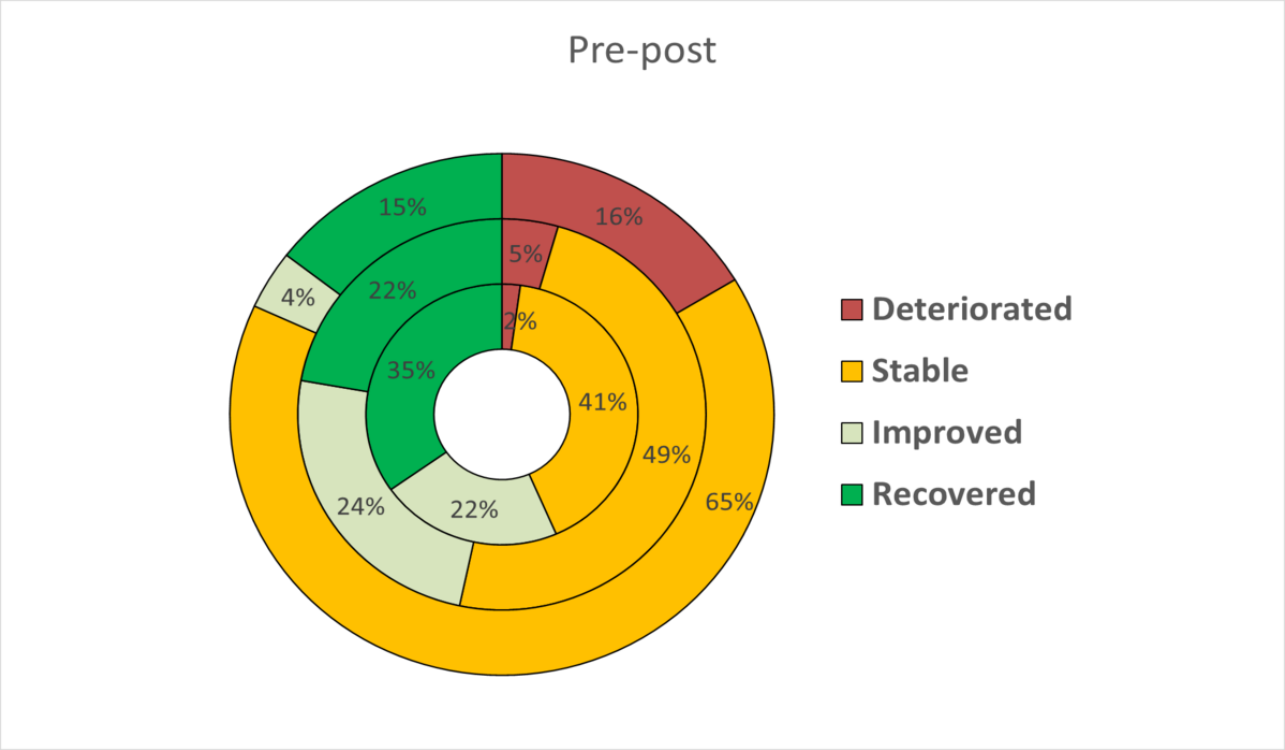
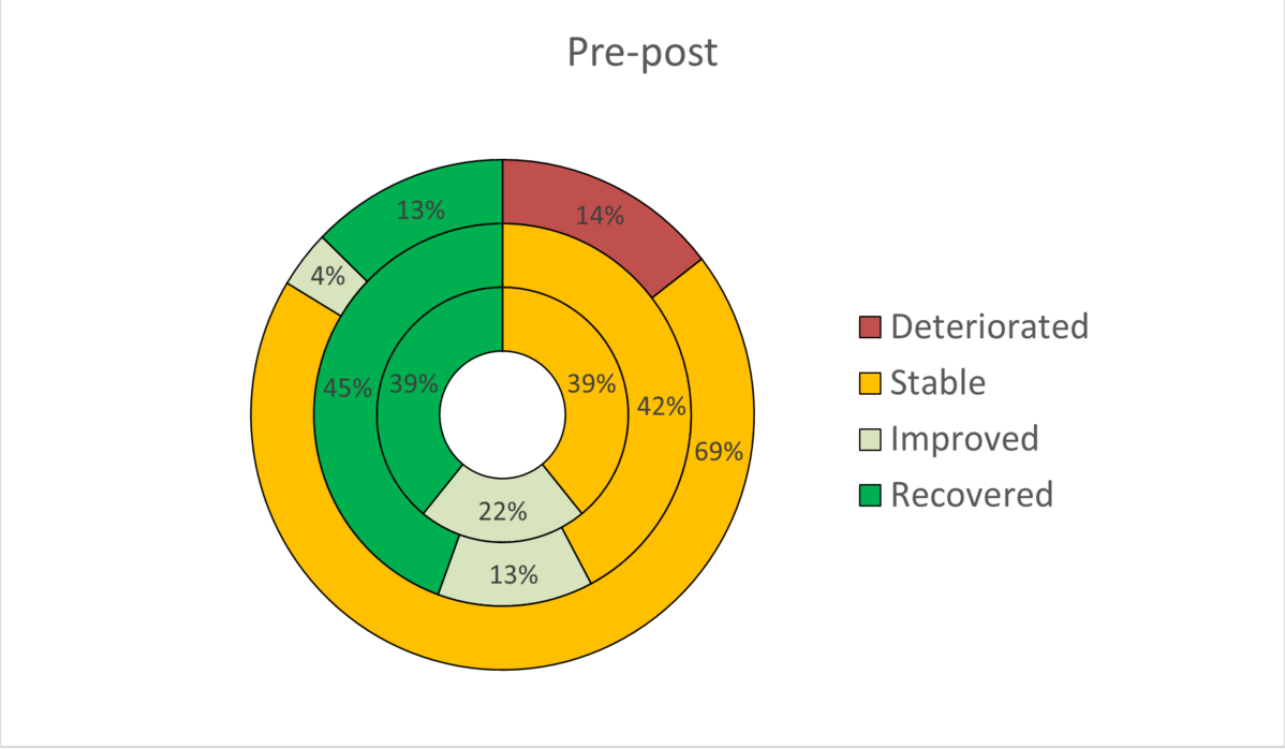


Figure 5. Percentages of participants recovered, improved, stable, and deteriorated on depression scores (Positive and Negative Affect Schedule_Negative subscale) in transdiagnostic internet-based protocol+positive affect component (inner circle), transdiagnostic internet-based protocol (middle circle), and waiting list (outer circle).



Discussion

Principal Findings

The main objective of this study was to test the efficacy of a transdiagnostic internet-based psychological treatment protocol, with and without a specific component to upregulate PA, versus a WL control group, for individuals from a community sample with one or more ED diagnoses. The results showed that on the primary outcome measures, there was a significant time effect, with medium (BAI and PANAS-P) to large effect sizes (BDI-II and PANAS-N) in the TIBP condition and large effect sizes on all measures in the TIBP+PA condition. In the WL group, the effect size was minimal. These results are comparable with those reported in other transdiagnostic trials in which effect sizes are larger for depression than for anxiety and higher for PANAS-P than for PANAS-N [13]. A possible explanation for the larger effects for depression than for anxiety could be that the therapeutic techniques more related to the treatment of anxiety disorders (ie, consequences of emotional avoidance, emotion-driven behaviors, and interoceptive and situational exposure) are presented near the end of the treatment protocol (modules from 8 to 11). Patients are evaluated clinically shortly after these modules, and consequently, they have less time to practice them. Likewise, the effects for PA were larger in the TIBP+PA condition (d=1.27) than in the TIBP condition (d=0.69). This result is consistent with the fact that this study includes a specific component to upregulate PA, which places greater emphasis on positive experiences. One of the aims of the intervention is to teach patients the adaptive role of both positive and negative emotions, without the need to suppress or avoid them. Nevertheless, future studies based on mediational analyses will allow a better understanding of whether PA and NA would mediate the impact of treatment on therapy outcomes (ie, depression and anxiety). As predicted, both protocol modalities led to significantly greater reductions (relative to the WL) in depression and anxiety, as well as significant decreases in NA and increases in PA measures, suggesting that web-based transdiagnostic treatment was effective in treating EDs.

On the secondary outcomes (NEO FFI and EuroQoL-5D questionnaire, EQ-5D), the analysis also revealed a significant change from pre- to posttreatment. The effect sizes were medium for both variables in both intervention groups (except for Extraversion, which was smaller in the TIBP condition). In the WL group, small effect sizes were also found for Extraversion, showing that participants in this group experienced significantly lower levels of extraversion after the waiting period. Between-group comparisons revealed that participants who received the treatment scored better on the measures at posttreatment compared with the WL group. These findings contribute to the study of the *malleability* of neuroticism proposed by Barlow [86], who suggests that personality dimensions may be malleable over time. In this study, some strategies were proposed to modify PA (extraversion), and although significant effects were found between the two intervention groups and the WL group, further research is needed on this topic to explore the effect of this specific component on increasing well-being and PA. Along these lines, Barlow's team conducted an RCT that evaluated changes in PA in cognitive-behavioral treatments for anxiety disorders, showing that PA is a malleable construct and can be influenced by CBT [87].

The second aim of this study was to assess the effects of the component designed to upregulate PA, with the hypothesis that the TIBP+PA condition would significantly outperform the TIBP condition on the PA measures. This study found no significant differences between the two treatment conditions on PA measures. This may be due to the fact that participants improved throughout the course of the treatment, decreasing negative symptoms and increasing their PA, even if they had not seen the last four modules with the specific component to upregulate PA. Therefore, future studies will be needed to reach a deeper understanding of the relationship between treatments that include PA components and changes in PA measures.

Another objective of the study was to study patients' acceptance of the program developed to apply the treatment protocol over the internet with minimal support from clinicians. The overall expectations and satisfaction expressed by the participants in the two intervention groups were high, and both protocols were well accepted. However, the analysis revealed higher levels of expectations and satisfaction in the TIBP+PA condition. This may be due to the number of modules included in this condition (16) in comparison with the modules in the TIBP condition (12), although participants in the intervention groups did not know the content of those four additional modules. This result may be influenced by the role of expectations before and during a web-based psychological intervention, which is consistent with the literature showing that treatment expectations became more favorable over time [88]. Regarding the reliable change index, significant improvements were found in the two treatment conditions compared with the WL group. Overall, participants who received the transdiagnostic internet-based interventions showed higher recovery percentages and less deterioration compared with the WL group.

This study had a dropout rate of 38.9% (86/216). Some studies in the literature have indicated dropout rates of approximately 30% in computerized CBT programs [47]. However, this issue deserves special attention because it depends on how it is conceptualized. On the one hand, the definition of attrition differs and can be understood as premature termination [89], (non)persistence [90], (non)adherence, or the extent to which an individual is exposed to the content of an intervention [91]. In this study, only participants who completed all the treatment modules were considered completers, whereas the rest were treated as dropouts, although in other studies, only participants who completed a percentage of the modules were considered completers [92,93]. On the other hand, adherence to internet-based interventions is associated with the type of guidance, achieving 28% nonadherence when therapist-guided, and 38% when administrative support is given [94]. This study combined human and ICT support, reaching a 36.8% (53/144) dropout rate in the treatment conditions; therefore, these results are not far from those found in the existing literature on internet-based interventions. However, further research on attrition is needed to better design internet-based treatments and increase retention. The existing literature has indicated an important relationship between the support provided to the

participants and adherence to the program. Indeed, this aspect has great relevance in the literature on internet-based interventions [95,96]. Therefore, if this type of intervention is carried out on a large scale in the community, as in this study, a fundamental factor to consider is the type and amount of support to provide, which would be consistent with the study of the balance between the benefits and resources involved in providing human support in internet-based interventions. Technologies play a central role in this aspect because part of the support can be provided automatically through technological devices. In an attempt to understand this issue, our research group conducted a qualitative study to determine why patients dropped out of the web-based transdiagnostic program for EDs (TIBP) [97], with the results emphasizing the lack of individualization of the treatment or the lack of support (ie, the lack of affective and personal contact with the therapist). In this regard, future studies should develop personalized treatments to address patients' specific needs and increase adherence rates. There are some useful strategies for personalizing treatments, such as selecting certain treatment components to better fit the patient's symptoms or lowering the number of sessions required to successfully treat an individual's symptoms. One strategy implies personalizing the treatment to a specific presentation, that is, by selecting the treatment components that best fit the specific set of symptoms or *weaknesses* shown by each patient [10], thereby lowering the number of sessions required to successfully treat an individual's symptoms. Another example is the study by Carl et al [98]. In this study, the authors presented a module for the regulation of PA that can be added to the treatment once they have completed the UP modules, thus personalizing the treatment for patients who show deficits in positive emotions at posttreatment.

This study has several strengths. First, it presents a novel focus in the field of transdiagnostic treatments. To the best of our knowledge, this is the first study of a transdiagnostic internet-based treatment for EDs with a specific component to upregulate PA. Overall, the findings indicate that EDs can be effectively treated with a transdiagnostic intervention via the internet, in addition to improving depression, anxiety, and quality of life measures. Regarding PA measures, promising effects were found, but more research is needed to study the role of positive emotions in the construction of psychological strengths [99,100] from a transdiagnostic perspective [42,101]. Moreover, this study included a large sample of people from a community sample, representing a heterogeneous population with EDs that does not receive primary or specialized care, with or without the presence of comorbidities. Thus, the transdiagnostic protocol represents a successful approach to the treatment of multiple disorders in a parsimonious manner [102]. Participants in the study seemed to be interested in the use of adaptive emotion regulation strategies, regardless of whether they were related to their own difficulty, which is the basis of transdiagnostic proposals. These interventions emphasize the essential processes underlying different disorders and the use of core *higher-order* strategies that eliminate the need for multiple diagnosis-specific manuals [103]. Moreover, the internet-based format of this transdiagnostic protocol facilitates the availability and administration of the program to provide support to anyone in need.

Limitations

This study also has some limitations. The most important is the different number of modules in the two protocols (TIBP condition: 12 modules; TIBP+PA condition: 16 modules). The TIBP condition had more time between the last module and postassessment than the TIBP+PA condition. However, the TIBP+PA condition had more modules. Postassessment took place at the end of module 12 or 16. These aspects may have influenced the results. In an attempt to control this, equal time (ie, a maximum period of 18 weeks from randomization to posttreatment evaluation) was given to all participants to allow them to use the program as much as they desired throughout the whole process. However, future studies should show that the differential effect of the PA component is not simply because of the larger number of modules in the protocol. This leads to the importance of benchmarking based on previous transdiagnostic internet-based interventions with regard to effect sizes or the length of these interventions. This study coincides with previous similar transdiagnostic interventions that also obtained large effect sizes for depression (g=0.84) and medium effect sizes for anxiety (g=0.78), with a treatment length ranging from 6 to 10 sessions [18]. Furthermore, focusing on PA measures, it is important to mention that this study obtained larger effect sizes for PA, both for within- and between-comparisons (TIBP+PA, d=1.27; TIBP+PA vs WL, d=0.90) than other transdiagnostic interventions [13,14,50-53]. Another limitation of the study, shared with other transdiagnostic interventions, is that although they are called transdiagnostic treatments, they are based on discrete diagnostic categories (ie, DSM-IV). Future research should study the mechanistically transdiagnostic principles, that is, the underlying mechanisms that account for the occurrence of specific symptoms to include them in both assessments and transdiagnostic interventions. Some examples of transdiagnostic mechanisms that have been found to play a fundamental role in EDs are intolerance to uncertainty [104], rumination [105], perfectionism [106], or thought suppression [107]. Among these processes, neuroticism has been strongly associated with both anxiety and depressive disorders [86,108]. In addition, attrition rates were higher in both treatment conditions (TIBP: 26/71, 37%; TIBP+PA: 27/73, 37%) than in the WL condition (17/72, 24%). However, attrition rates of 30% to 35% are commonly observed in internet-based interventions [47]. Furthermore, although we consider that the sample of this study is quite representative, future research should focus on improving the mental health of less reachable groups such as older people, people with low income or lower educational level, and/or people residing outside urban areas (ie, rural populations). Moreover, it would also be important to collect data on certain demographic variables such as race, ethnicity, and sexual orientation in future trials. Transdiagnostic treatments that are socioculturally adapted for such groups could be a potential solution to effectively reach a larger number of people in need of psychological help.

Finally, follow-up data were not included in this study because we are still in the recruiting process, and we wanted to promptly provide the results because of the importance of the issue. The long-term effects of this intervention will be presented in further research.

Conclusions

In summary, the results show the efficacy of a transdiagnostic internet-based psychological treatment protocol for individuals from a community sample with EDs. These findings show that this web-based transdiagnostic treatment improved the clinical situation of the participants, providing them with tools and strategies to face problems and difficulties more effectively. Future lines of research should carry out dismantling designs to determine the active components of the protocol, especially the contribution of the PA modules, and analyze the effectiveness of web-based treatment in other populations, such as primary care centers. Furthermore, the existing techniques and strategies to improve PA require further study to determine which ones are more effective and should be included as specific components to upregulate PA in current psychological interventions. This study includes some specific strategies to promote psychological strengths and enhance positive mood. However, it is important for future studies to explore more deeply the effect of these and other strategies on patients with EDs to directly build positive resources to counteract NA. Undoubtedly, future research will have to determine whether it is beneficial to include these components designed to enhance PA, which components are necessary for whom, and how they should be applied.

Furthermore, future research should focus on the possibility of developing treatment components aimed at altering, modifying, or varying vulnerability, a key aspect of transdiagnostic perspectives. This might be possible with strategies for modifying PA, but there are other fundamental factors that influence mental health and well-being. For example, a large body of literature has highlighted the importance of accurate perceptions of reality in psychological health, such as positive illusions or positive self-evaluations, perceptions of control or mastery, and unrealistic optimism [109,110]. In addition, there is evidence for the relationship between psychological flexibility and well-being, suggesting that being psychologically flexible has benefits for executive functioning, default mental states, and personality dimensions such as neuroticism [111]. Psychological flexibility has been considered a protective factor in improving physical health, mental health, and well-being [112]. Hence, there is growing interest in constructs such as pragmatic prospection, that is, thinking about a future with desired outcomes and avoiding undesired ones [113] or openness to the future characterized by PA toward the future [114]. This body of knowledge opens up the possibility of finding new strategies to improve the efficiency and effectiveness of future transdiagnostic treatment protocols for EDs as a way to more effectively address temperament vulnerabilities, that is, the core aspects of these disorders.

Acknowledgments

This study was funded by the Ministry of Economy and Competitiveness (Spain; PSI2014-54172-R), a PhD grant from the Ministry of Economy and Competitiveness (FPI-MINECO; BES-2015-072360), and CIBER Fisiopatología de la Obesidad y Nutrición-ISCIII CB06/03/0052.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of the positive affect modules.
[DOC File , 38 KB - jmir_v23i2e21335_app1.doc]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1638 KB - jmir_v23i2e21335_app2.pdf]

References

1. Bullis JR, Boettcher H, Sauer‐Zavala S, Farchione TJ, Barlow DH. What is an emotional disorder? A transdiagnostic mechanistic definition with implications for assessment, treatment, and prevention. Clin Psychol Sci Pract 2019 Mar 14;26(2):e12278. [doi: 10.1111/cpsp.12278]
2. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
3. Nathan P, Gorman J. A Guide to Treatments That Work. In: 4th Edition. United States of America: Oxford University Press; 2015.
4. Kazdin AE, Blase SL. Rebooting Psychotherapy Research and Practice to Reduce the Burden of Mental Illness. Perspectives on Psychological Science 2011 Feb 03;6(1):21-37 [FREE Full text] [doi: 10.1177/1745691610393527]
5. Ellard KK, Fairholme CP, Boisseau CL, Farchione TJ, Barlow DH. Unified Protocol for the Transdiagnostic Treatment of Emotional Disorders: Protocol Development and Initial Outcome Data. Cognitive and Behavioral Practice 2010 Feb;17(1):88-101. [doi: 10.1016/j.cbpra.2009.06.002]

6. McHugh R, Barlow D. The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. Am Psychol 2010;65(2):73-84 [FREE Full text] [doi: 10.1037/a0018121]
7. Barlow DH, Allen LB, Choate ML. Toward a unified treatment for emotional disorders. Behavior Therapy 2004 Mar;35(2):205-230. [doi: 10.1016/S0005-7894(04)80036-4]
8. Brown TA, Barlow DH. A proposal for a dimensional classification system based on the shared features of the DSM-IV anxiety and mood disorders: Implications for assessment and treatment. Psychological Assessment 2009;21(3):256-271. [doi: 10.1037/a0016608]
9. Aldao A, Nolen-Hoeksema S, Schweizer S. Emotion-regulation strategies across psychopathology: A meta-analytic review. Clinical Psychology Review 2010 Mar;30(2):217-237. [doi: 10.1016/j.cpr.2009.11.004]
10. Sauer-Zavala S, Cassiello-Robbins C, Ametaj AA, Wilner JG, Pagan D. Transdiagnostic Treatment Personalization: The Feasibility of Ordering Unified Protocol Modules According to Patient Strengths and Weaknesses. Behav Modif 2018 May 10;43(4):518-543. [doi: 10.1177/0145445518774914]
11. Norton PJ, Paulus DJ. Toward a Unified Treatment for Emotional Disorders: Update on the Science and Practice. Behav Ther 2016 Nov;47(6):854-868. [doi: 10.1016/j.beth.2015.07.002] [Medline: 27993337]
12. Barlow DH, Farchione TJ, Bullis JR, Gallagher MW, Murray-Latin H, Sauer-Zavala S, et al. The Unified Protocol for Transdiagnostic Treatment of Emotional Disorders Compared With Diagnosis-Specific Protocols for Anxiety Disorders: A Randomized Clinical Trial. JAMA Psychiatry 2017 Aug 02:-. [doi: 10.1001/jamapsychiatry.2017.2164] [Medline: 28768327]
13. Farchione TJ, Fairholme CP, Ellard KK, Boisseau CL, Thompson-Hollands J, Carl JR, et al. Unified Protocol for Transdiagnostic Treatment of Emotional Disorders: A Randomized Controlled Trial. Behavior Therapy 2012 Sep;43(3):666-678. [doi: 10.1016/j.beth.2012.01.001]
14. Bullis JR, Fortune MR, Farchione TJ, Barlow DH. A preliminary investigation of the long-term outcome of the Unified Protocol for Transdiagnostic Treatment of Emotional Disorders. Comprehensive Psychiatry 2014 Nov;55(8):1920-1927. [doi: 10.1016/j.comppsych.2014.07.016]
15. Carl JR, Gallagher MW, Sauer-Zavala SE, Bentley KH, Barlow DH. A preliminary investigation of the effects of the unified protocol on temperament. Comprehensive Psychiatry 2014 Aug;55(6):1426-1434. [doi: 10.1016/j.comppsych.2014.04.015]
16. Reinholt N, Krogh J. Efficacy of Transdiagnostic Cognitive Behaviour Therapy for Anxiety Disorders: A Systematic Review and Meta-Analysis of Published Outcome Studies. Cognitive Behaviour Therapy 2014 Mar 19;43(3):171-184. [doi: 10.1080/16506073.2014.897367]
17. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev 2015 Aug;40:91-110 [FREE Full text] [doi: 10.1016/j.cpr.2015.06.002] [Medline: 26094079]
18. Newby JM, Twomey C, Yuan LSS, Andrews G. Transdiagnostic computerised cognitive behavioural therapy for depression and anxiety: A systematic review and meta-analysis. J Affect Disord 2016 Jul 15;199:30-41. [doi: 10.1016/j.jad.2016.03.018] [Medline: 27060430]
19. Păsărelu CR, Andersson G, Bergman NL, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
20. Pearl SB, Norton PJ. Transdiagnostic versus diagnosis specific cognitive behavioural therapies for anxiety: A meta-analysis. J Anxiety Disord 2017 Mar;46:11-24. [doi: 10.1016/j.janxdis.2016.07.004] [Medline: 27466074]
21. Clark DA, Taylor S. The Transdiagnostic Perspective on Cognitive-Behavioral Therapy for Anxiety and Depression: New Wine for Old Wineskins? J Cogn Psychother 2009 Feb 01;23(1):60-66. [doi: 10.1891/0889-8391.23.1.60]
22. Sakiris N, Berle D. A systematic review and meta-analysis of the Unified Protocol as a transdiagnostic emotion regulation based intervention. Clinical Psychology Review 2019 Aug;72:101751. [doi: 10.1016/j.cpr.2019.101751]
23. Weinstock LM, Whisman MA. Neuroticism as a common feature of the depressive and anxiety disorders: A test of the revised integrative hierarchical model in a national sample. J Abnorm Psychol 2006;115(1):68-74. [doi: 10.1037/0021-843x.115.1.68]
24. Clark LA, Watson D, Mineka S. Temperament, personality, and the mood and anxiety disorders. J Abnorm Psychol 1994;103(1):103-116. [doi: 10.1037/0021-843x.103.1.103]
25. Kasch KL, Rottenberg J, Arnow BA, Gotlib IH. Behavioral activation and inhibition systems and the severity and course of depression. J Abnorm Psychol 2002;111(4):589-597. [doi: 10.1037/0021-843x.111.4.589]
26. Campbell-Sills L, Liverant GI, Brown TA. Psychometric evaluation of the behavioral inhibition/behavioral activation scales in a large sample of outpatients with anxiety and mood disorders. Psychol Assess 2004;16(3):244-254. [doi: 10.1037/1040-3590.16.3.244]
27. Brown TA. Temporal course and structural relationships among dimensions of temperament and DSM-IV anxiety and mood disorder constructs. J Abnorm Psychol 2007 May;116(2):313-328. [doi: 10.1037/0021-843X.116.2.313] [Medline: 17516764]
28. Campbell-Sills L, Barlow DH, Brown TA, Hofmann SG. Effects of suppression and acceptance on emotional responses of individuals with anxiety and mood disorders. Behav Res Ther 2006 Sep;44(9):1251-1263. [doi: 10.1016/j.brat.2005.10.001]

29. Kotov R, Gamez W, Schmidt F, Watson D. Linking 'big' personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. Psychol Bull 2010;136(5):768-821. [doi: 10.1037/a0020327]
30. Lopez S, Snyder C. The Oxford Handbook of Positive Psychology. New York: Oxford University Press; 2009.
31. Gilbert KE, Nolen-Hoeksema S, Gruber J. Positive emotion dysregulation across mood disorders: how amplifying versus dampening predicts emotional reactivity and illness course. Behav Res Ther 2013 Nov;51(11):736-741. [doi: 10.1016/j.brat.2013.08.004] [Medline: 24076407]
32. Sauer-Zavala S, Gutner CA, Farchione TJ, Boettcher HT, Bullis JR, Barlow DH. Current Definitions of "Transdiagnostic" in Treatment Development: A Search for Consensus. Behav Ther 2017 Jan;48(1):128-138. [doi: 10.1016/j.beth.2016.09.004] [Medline: 28077216]
33. Fredrickson BL. The broaden-and-build theory of positive emotions. Philos Trans R Soc Lond B Biol Sci 2004 Sep 29;359(1449):1367-1378 [FREE Full text] [doi: 10.1098/rstb.2004.1512] [Medline: 15347528]
34. Dockray S, Steptoe A. Positive affect and psychobiological processes. Neurosci Biobehav Rev 2010 Sep;35(1):69-75 [FREE Full text] [doi: 10.1016/j.neubiorev.2010.01.006] [Medline: 20097225]
35. Tugade MM, Fredrickson BL. Resilient individuals use positive emotions to bounce back from negative emotional experiences. J Pers Soc Psychol 2004 Feb;86(2):320-333 [FREE Full text] [doi: 10.1037/0022-3514.86.2.320] [Medline: 14769087]
36. Sewart AR, Zbozinek TD, Hammen C, Zinbarg RE, Mineka S, Craske MG. Positive Affect as a Buffer between Chronic Stress and Symptom Severity of Emotional Disorders. Clin Psychol Sci 2019 Sep;7(5):914-927 [FREE Full text] [doi: 10.1177/2167702619834576] [Medline: 31632843]
37. Pressman SD, Jenkins BN, Moskowitz JT. Positive Affect and Health: What Do We Know and Where Next Should We Go? Annu Rev Psychol 2019 Jan 04;70:627-650. [doi: 10.1146/annurev-psych-010418-102955] [Medline: 30260746]
38. Blanco-Donoso LM, Moreno-Jiménez B, Rodríguez-Carvajal R, Garrosa E. Regulación emocional y afecto positivo: el efecto mediador del sentido de la coherencia. Rev Argen Clin Psic 2018;27(3):403-412. [doi: 10.24205/03276716.2018.1079]
39. Bolier L, Haverman M, Westerhof GJ, Riper H, Smit F, Bohlmeijer E. Positive psychology interventions: a meta-analysis of randomized controlled studies. BMC Public Health 2013 Feb 08;13:119 [FREE Full text] [doi: 10.1186/1471-2458-13-119] [Medline: 23390882]
40. Roepke AM, Seligman MEP. Depression and prospection. Br J Clin Psychol 2015 Jun 12;55(1):23-48. [doi: 10.1111/bjc.12087]
41. Titov N, Dear BF, Schwencke G, Andrews G, Johnston L, Craske MG, et al. Transdiagnostic internet treatment for anxiety and depression: a randomised controlled trial. Behav Res Ther 2011 Aug;49(8):441-452. [doi: 10.1016/j.brat.2011.03.007] [Medline: 21679925]
42. Carl JR, Soskin DP, Kerns C, Barlow DH. Positive emotion regulation in emotional disorders: A theoretical review. Clinical Psychology Review 2013 Apr;33(3):343-360. [doi: 10.1016/j.cpr.2013.01.003]
43. Craske MG, Meuret AE, Ritz T, Treanor M, Dour H, Rosenfield D. Positive affect treatment for depression and anxiety: A randomized clinical trial for a core feature of anhedonia. J Consult Clin Psychol 2019 May;87(5):457-471. [doi: 10.1037/ccp0000396] [Medline: 30998048]
44. Kazdin AE. Technology-Based Interventions and Reducing the Burdens of Mental Illness: Perspectives and Comments on the Special Series. Cognitive and Behavioral Practice 2015 Aug;22(3):359-366. [doi: 10.1016/j.cbpra.2015.04.004]
45. Andersson G. Internet-Delivered Psychological Treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
46. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
47. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78. [doi: 10.1016/j.janxdis.2018.01.001]
48. Cuijpers P, van SA, Warmerdam L, Andersson G. Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. Depress Anxiety 2009;26(3):279-288. [doi: 10.1002/da.20519] [Medline: 19031487]
49. American Psychological Association APA. Diagnostic And Statistical Manual Of Mental Disorders. 4th ed. Washington DC: American Psychiatric Association; 2000:A.
50. Ito M, Horikoshi M, Kato N, Oe Y, Fujisato H, Nakajima S, et al. Transdiagnostic and Transcultural: Pilot Study of Unified Protocol for Depressive and Anxiety Disorders in Japan. Behav Ther 2016 May;47(3):416-430. [doi: 10.1016/j.beth.2016.02.005] [Medline: 27157034]
51. Osma J, Castellano C, Crespo E, Garcia-Palacios A. The unified protocol for transdiagnostic treatment of emotional disorders in group format in a Spanish public mental health setting. Behav Psychol/Psicol Conductual 2015;23(3):447-466. [doi: 10.1093/med-psych/9780190255541.003.0001]
52. Reinholt N, Aharoni R, Winding C, Rosenberg N, Rosenbaum B, Arnfred S. Transdiagnostic group CBT for anxiety disorders: the unified protocol in mental health services. Cogn Behav Ther 2017 Jan;46(1):29-43. [doi: 10.1080/16506073.2016.1227360] [Medline: 27705086]

53. Taylor CT, Lyubomirsky S, Stein MB. Upregulating the positive affect system in anxiety and depression: Outcomes of a positive activity intervention. Depress Anxiety 2017 Mar;34(3):267-280 [FREE Full text] [doi: 10.1002/da.22593] [Medline: 28060463]
54. Moher D, Hopewell S, Schulz K, Montori V, Gøtzsche PC, Devereaux PJ, CONSORT. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Int J Surg 2012;10(1):28-55 [FREE Full text] [doi: 10.1016/j.ijsu.2011.10.001] [Medline: 22036893]
55. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
56. Chan A, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013 Feb 05;158(3):200-207 [FREE Full text] [doi: 10.7326/0003-4819-158-3-201302050-00583] [Medline: 23295957]
57. Chan A, Tetzlaff JM, Gotzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. Br Med J 2013 Jan 09;346(jan08 15):e7586-e7586. [doi: 10.1136/bmj.e7586]
58. Díaz-García A, González-Robles A, Fernández-Álvarez J, García-Palacios A, Baños RM, Botella C. Efficacy of a Transdiagnostic internet-based treatment for emotional disorders with a specific component to address positive affect: Study protocol for a randomized controlled trial. BMC Psychiatry 2017 Apr 20;17(1):-. [doi: 10.1186/s12888-017-1297-z]
59. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/BF03193146]
60. Ferrando L, Bobes J, Gibert J, Lecubrier Y. Mini international neuropsychiatric interview. In: Manejo De Los Trastornos Mentales Y Del Comportamiento en Atención Primaria. Oviedo: Gofer; 1997.
61. Linehan M. Cognitive-Behavioral Treatment Of Borderline Personality Disorder. New York: The Guilford Press; 2010.
62. Labpsitec. 2011. Psicología y Tecnología https://www.psicologiaytecnologia.com/. URL: https://www.psicologiaytecnologia.com/ [accessed 2015-09-10]
63. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009 May;65(5):467-487. [doi: 10.1002/jclp.20593] [Medline: 19301241]
64. Lejuez CW, Hopko DR, Hopko SD. A brief behavioral activation treatment for depression. Treatment manual. Behav Modif 2001 Apr;25(2):255-286. [doi: 10.1177/0145445501252005] [Medline: 11317637]
65. Seligman ME, Csikszentmihalyi M. Positive psychology: an introduction. Am Psychol 2000;55(1):5-14. [doi: 10.1037/0003-066X.55.1.5]
66. Fava GA. Well-Being Therapy: conceptual and technical issues. Psychother Psychosom 1999;68(4):171-179. [doi: 10.1159/000012329]
67. Fava GA, Ruini C. Development and characteristics of a well-being enhancing psychotherapeutic strategy: well-being therapy. J Behav Ther Exp Psychiatry 2003 Mar;34(1):45-63. [doi: 10.1016/s0005-7916(03)00019-3]
68. Fredrickson BL. The role of positive emotions in positive psychology: the broaden-and-build theory of positive emotions. Am Psychol 2001;56(3):218-226. [doi: 10.1037/0003-066x.56.3.218]
69. Ekman P, Davidson RJ, Friesen WV. The Duchenne smile: emotional expression and brain physiology. II. J Pers Soc Psychol 1990 Feb;58(2):342-353. [Medline: 2319446]
70. Soussignan R. Duchenne smile, emotional experience, and autonomic reactivity: a test of the facial feedback hypothesis. Emotion 2002 Mar;2(1):52-74. [doi: 10.1037/1528-3542.2.1.52] [Medline: 12899366]
71. Miles L, Johnston L. Detecting Happiness: Perceiver Sensitivity to Enjoyment and Non-Enjoyment Smiles. J Nonverbal Behav 2007 Sep 25;31(4):259-275. [doi: 10.1007/s10919-007-0036-4]
72. Bryant FB. Savoring: a New Model of Positive Experience. Mahwah, NJ: Erlbaum; 2007.
73. Peterson C. Character Strengths and Virtues: a Handbook and Classification. Oxford, UK: Oxford University Press; 2004.
74. Ryff CD, Keyes CL. The structure of psychological well-being revisited. J Pers Soc Psychol 1995 Oct;69(4):719-727. [doi: 10.1037//0022-3514.69.4.719] [Medline: 7473027]
75. Ryff CD. Psychological well-being revisited: advances in the science and practice of eudaimonia. Psychother Psychosom 2014;83(1):10-28 [FREE Full text] [doi: 10.1159/000353263] [Medline: 24281296]
76. Seligman ME, Steen TA, Park N, Peterson C. Positive psychology progress: empirical validation of interventions. Am Psychol 2005;60(5):410-421. [doi: 10.1037/0003-066X.60.5.410] [Medline: 16045394]
77. Sheldon KM, Lyubomirsky S. How to increase and sustain positive emotion: the effects of expressing gratitude and visualizing best possible selves. J Positive Psychol 2006 Apr;1(2):73-82. [doi: 10.1080/17439760500510676]
78. Commify IS. 2019. Trendoo https://www.trendoo.es/. URL: https://www.trendoo.es/ [accessed 2015-09-15]
79. Chakraborty H. A mixed model approach for intent-to-treat analysis in longitudinal clinical trials with missing values. RTI Press Methods Report Series 2009:e. [doi: 10.3768/rtipress.2009.mr.0009.0903] [Medline: 30896910]
80. Hesser H. Modeling individual differences in randomized experiments using growth models: Recommendations for design, statistical analysis and reporting of results of internet interventions. Internet Interv 2015 May;2(2):110-120. [doi: 10.1016/j.invent.2015.02.003]

81. Salim A, Mackinnon A, Christensen H, Griffiths K. Comparison of data analysis strategies for intent-to-treat analysis in pre-test–post-test designs with substantial dropout rates. Psychiatry Res 2008 Sep;160(3):335-345. [doi: 10.1016/j.psychres.2007.08.005]
82. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: N.J: L. Erlbaum Associates; 1988.
83. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59(1):12-19. [doi: 10.1037/0022-006X.59.1.12]
84. Twisk J, de BM, de VW, Heymans M. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. J Clin Epidemiol 2013 Sep;66(9):1022-1028. [doi: 10.1016/j.jclinepi.2013.03.017] [Medline: 23790725]
85. Peters SA, Bots ML, den RH, Palmer MK, Grobbee DE, Crouse JR, et al. Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models. J Clin Epidemiol 2012 Jun;65(6):686-695. [doi: 10.1016/j.jclinepi.2011.11.012] [Medline: 22459429]
86. Barlow DH, Sauer-Zavala S, Carl JR, Bullis JR, Ellard KK. The nature, diagnosis, and treatment of neuroticism. Clin Psychol Sci 2013 Oct 14;2(3):344-365. [doi: 10.1177/2167702613505532]
87. Wilner Tirpak J, Cassiello-Robbins C, Ametaj A, Olesnycky OS, Sauer-Zavala S, Farchione TJ, et al. Changes in positive affect in cognitive-behavioral treatment of anxiety disorders. Gen Hosp Psychiatry 2019 Nov;61:111-115. [doi: 10.1016/j.genhosppsych.2019.06.008]
88. Meyerhoff J, Rohan KJ. Treatment expectations for cognitive-behavioral therapy and light therapy for seasonal affective disorder: Change across treatment and relation to outcome. J Consult Clin Psychol 2016 Oct;84(10):898-906. [doi: 10.1037/ccp0000121]
89. Hatchett GT, Park HL. Comparison of four operational definitions of premature termination. Psychother Theo Res Pract Train 2003;40(3):226-231. [doi: 10.1037/0033-3204.40.3.226]
90. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012 Jun;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
91. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
92. Erickson DH. Group cognitive behavioural therapy for heterogeneous anxiety disorders. Cogn Behav Ther 2003 Nov;32(4):179-186. [doi: 10.1080/16506070310001686]
93. García-Escalera J, Chorot P, Sandín B, Ehrenreich-May J, Prieto A, Valiente RM. An open trial applying the unified protocol for transdiagnostic treatment of emotional disorders in adolescents (UP-A) adapted as a school-based prevention program. Child Youth Care Forum 2018 Sep 12;48(1):29-53. [doi: 10.1007/s10566-018-9471-0]
94. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
95. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
96. Mira A, Bretón-López J, García-Palacios A, Quero S, Baños RM, Botella C. An Internet-based program for depressive symptoms using human and automated support: a randomized controlled trial. Neuropsychiatr Dis Treat 2017 Mar;Volume 13:987-1006. [doi: 10.2147/ndt.s130994]
97. Fernández-Álvarez J, Díaz-García A, González-Robles A, Baños R, García-Palacios A, Botella C. Dropping out of a transdiagnostic online intervention: a qualitative analysis of client's experiences. Internet Interv 2017 Dec;10:29-38. [doi: 10.1016/j.invent.2017.09.001]
98. Carl JR, Gallagher MW, Barlow DH. Development and preliminary evaluation of a positive emotion regulation augmentation module for anxiety and depression. Behav Ther 2018 Nov;49(6):939-950. [doi: 10.1016/j.beth.2017.11.008]
99. Algoe SB, Fredrickson BL. Emotional fitness and the movement of affective science from lab to field. Am Psychol 2011;66(1):35-42. [doi: 10.1037/a0021720]
100. Catalino LI, Fredrickson BL. A Tuesday in the life of a flourisher: the role of positive emotional reactivity in optimal mental health. Emotion 2011;11(4):938-950. [doi: 10.1037/a0024889]
101. Carl JR, Fairholme CP, Gallagher MW, Thompson-Hollands J, Barlow DH. The effects of anxiety and depressive symptoms on daily positive emotion regulation. J Psychopathol Behav Assess 2013 Sep 24;36(2):224-236. [doi: 10.1007/s10862-013-9387-9]
102. Barlow DH, Farchione TJ, Bullis JR, Gallagher MW, Murray-Latin H, Sauer-Zavala S, et al. The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders: a randomized clinical trial. JAMA Psychiatry 2017 Aug 02:-. [doi: 10.1001/jamapsychiatry.2017.2164] [Medline: 28768327]
103. Mansell W, Harvey A, Watkins E, Shafran R. Conceptual foundations of the transdiagnostic approach to CBT. J Cogn Psychother 2009 Feb 01;23(1):6-19. [doi: 10.1891/0889-8391.23.1.6]
104. Mahoney AE, McEvoy PM. A transdiagnostic examination of intolerance of uncertainty across anxiety and depressive disorders. Cogn Behav Ther 2012 Sep;41(3):212-222. [doi: 10.1080/16506073.2011.622130]
105. Ehring T, Watkins ER. Repetitive negative thinking as a transdiagnostic process. Int J Cogn Ther) 2008 Sep;1(3):192-205. [doi: 10.1521/ijct.2008.1.3.192]

106. Egan SJ, Wade TD, Shafran R. Perfectionism as a transdiagnostic process: a clinical review. Clin Psychol Rev 2011 Mar;31(2):203-212. [doi: 10.1016/j.cpr.2010.04.009] [Medline: 20488598]
107. Aldao A, Nolen-Hoeksema S. Specificity of cognitive emotion regulation strategies: a transdiagnostic examination. Behav Res Ther 2010 Oct;48(10):974-983. [doi: 10.1016/j.brat.2010.06.002]
108. Harvey A, Watkins E, Mansell W, Shafran R. Cognitive behavioural processes across psychological disorders: A transdiagnostic approach to research and treatment. New York: NY: Oxford University Press; 2004:A.
109. Taylor SE, Brown JD. Illusion and well-being: a social psychological perspective on mental health. Psychol Bull 1988;103(2):193-210. [doi: 10.1037/0033-2909.103.2.193]
110. Taylor SE, Kemeny ME, Reed GM, Bower JE, Gruenewald TL. Psychological resources, positive illusions, and health. Am Psychol 2000;55(1):99-109. [doi: 10.1037/0003-066x.55.1.99]
111. Kashdan TB, Rottenberg J. Psychological flexibility as a fundamental aspect of health. Clin Psychol Rev 2010 Nov;30(7):865-878 [FREE Full text] [doi: 10.1016/j.cpr.2010.03.001] [Medline: 21151705]
112. Gloster AT, Meyer AH, Lieb R. Psychological flexibility as a malleable public health target: evidence from a representative sample. J Contextual Behav Sci 2017 Apr;6(2):166-171. [doi: 10.1016/j.jcbs.2017.02.003]
113. Baumeister RF, Vohs KD, Oettingen G. Pragmatic prospection: how and why people think about the future. Rev Gen Psychol 2016 Mar;20(1):3-16. [doi: 10.1037/gpr0000060]
114. Botella C, Molinari G, Fernández-Álvarez J, Guillén V, García-Palacios A, Baños RM, et al. Development and validation of the openness to the future scale: a prospective protective factor. Health Qual Life Outcomes 2018 Apr 23;16(1):-. [doi: 10.1186/s12955-018-0889-8]

Abbreviations

AG: agoraphobia
BA: behavioral activation
BAI: Beck Anxiety Inventory
BDI-II: Beck Depression Inventory, Second Edition
BI: behavioral inhibition
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
DD: dysthymic disorder
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
E: extraversion
EBT: evidence-based psychological treatment
ED: emotional disorder
GAD: generalized anxiety disorder
ICT: information and communication technology
ITT: intention-to-treat
MDD: major depressive disorder
MNAR: missing not at random
N: neuroticism
NA: negative affect or negative affectivity
NEO FFI: NEO Five Factor Inventory
PA: positive affect
PANAS: Positive and Negative Affect Schedule
PD: panic disorder
RCI: reliable change index
RCT: randomized controlled trial
SAD: social anxiety disorder
TIBP: transdiagnostic internet-based protocol
UP: Unified Protocol
WL: waiting list

Edited by T Rashid Soron; submitted 11.06.20; peer-reviewed by L Eustis Ph.D, E Kelly, C Păsărelu; comments to author 23.07.20; revised version received 10.09.20; accepted 01.12.20; published 01.02.21.

Please cite as:
Díaz-García A, González-Robles A, García-Palacios A, Fernández-Álvarez J, Castilla D, Bretón JM, Baños RM, Quero S, Botella C Negative and Positive Affect Regulation in a Transdiagnostic Internet-Based Protocol for Emotional Disorders: Randomized Controlled Trial
J Med Internet Res 2021;23(2):e21335
URL: https://www.jmir.org/2021/2/e21335
doi:10.2196/21335
PMID:33522977

Original Paper

Two Web-Based and Theory-Based Interventions With and Without Brief Motivational Interviewing in the Promotion of Human Papillomavirus Vaccination Among Chinese Men Who Have Sex With Men: Randomized Controlled Trial

Zixin Wang[1], PhD; Joseph T F Lau[1], PhD; Tsun Kwan Mary Ip[1], MSc; Yebo Yu[1], MPH; Francois Fong[2], MD; Yuan Fang[3], PhD; Phoenix K H Mo[1], PhD

[1]JC School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China

[2]NeoHealth, Hong Kong, China

[3]Department of Early Childhood Education, Faculty of Education and Human Development, The Education University of Hong Kong, Hong Kong, China

Corresponding Author:
Zixin Wang, PhD
JC School of Public Health and Primary Care
Faculty of Medicine
The Chinese University of Hong Kong
Room 508, School of Public Health,
Prince of Wales Hospital, Shatin, NT
Hong Kong, 666888
China
Phone: 86 22528740
Email: wangzx@cuhk.edu.hk

Abstract

Background: Human papillomavirus (HPV) vaccination is effective in the prevention of vaccine-type genital warts and cancers among men who have sex with men (MSM).

Objective: The primary objective of this randomized controlled trial (RCT) is to evaluate the efficacies of 2 web- and theory–based interventions with and without brief motivational interviewing (MI) over the phone to increase the completion of HPV vaccination among unvaccinated participants within a 24-month follow-up period compared with the control group.

Methods: A 3-arm parallel-group RCT was conducted between July 2017 and December 2019. Five telephone surveys were conducted at baseline and at 3, 6, 9, and 24 months by blinded interviewers. Participants were Hong Kong Chinese–speaking MSM aged between 18 and 45 years with regular internet access who were recruited from outreaching at venues, web-based recruitment, and peer referral. Those who had ever received HPV vaccination were excluded. A total of 624 participants were randomized into either the online tutorial (OT) only group (n=208), the OT plus MI group (OT-MI; n=208), or the control group (n=208). In total, 459 (459/624, 73.6%) completed the follow-up evaluation at 24 months. Participants in the OT group received a fully automated OT developed based on the health belief model. On top of the same OT, the OT-MI group received brief MI over the phone. Reminders were sent to the participants of the OT and OT-MI groups after 1, 2, 4, 6, and 8 months. Participants in the control group received web-based health communication messages unrelated to HPV or HPV vaccination. The research team validated the self-reported HPV vaccination uptake. Intention-to-treat analysis was used for outcome analyses. Logistic regression models and multivariable linear regression models were used to test the between-group differences in primary and secondary outcomes. Baron and Kenny's methods were used to test the mediation hypothesis.

Results: The participants in the OT-MI group reported a significantly higher validated completion of HPV vaccination at 24 months than the control group (36/208, 17.3% vs 15/208, 7.2%; *P*=.006). However, the difference in HPV vaccination completion between the OT and the control groups (24/208, 11.5% vs 15/208, 7.2%; *P*=.17), or between OT-MI and OT groups (*P*=.13), was not statistically significant. The association between randomization status (OT-MI group vs control group) and HPV vaccination completion became statistically nonsignificant after controlling for changes in the perceived susceptibility to HPV (24 months

vs baseline), whereas perceived susceptibility remained strongly associated with HPV vaccination uptake in the model (P<.001). Changes in perceived susceptibility fully mediated the intervention effect.

Conclusions: Theory-based OT with brief MI over the phone was effective in increasing HPV vaccination completion among Chinese MSM. Perceived susceptibility is an active theoretical component that causes behavioral changes.

Trial Registration: ClinicalTrials.gov NCT03286907; https://clinicaltrials.gov/ct2/show/NCT03286907

(J Med Internet Res 2021;23(2):e21465) doi:10.2196/21465

KEYWORDS

HPV vaccination; web-based health promotion; randomized controlled trial; men who have sex with men; China; mobile phone

Introduction

Men who have sex with men (MSM) are at a high risk of contracting human papillomavirus (HPV) and its related diseases (eg, genital warts and penile or anal cancers) [1]. Meta-analyses have reported that the overall prevalence of genital HPV infection was very high among MSM both internationally (63.9% in HIV-negative MSM and 92.6% in HIV-infected MSM) [2] and in China (66.3% among MSM in general) [3]. This prevalence was much higher than in MSM than that in the general male population (eg, 12.4% in Europe and 16.9% in China) [4-6]. In addition to the high prevalence of genital warts [7], the anal cancer risk of MSM is 32 to 52 times higher than that in the general population [8]. The HPV-related cancer risk was the highest among HIV-infected MSM, which accounted for 9.9% of the MSM population in China in 2016 [9]. Although there is a lack of data about the prevalence or incidence of HPV-related diseases among MSM in Hong Kong, such prevalence may be much higher among MSM than that among the general male population (genital warts prevalence 0.94% and incidence 292.2 per 100,000 person-years) [10].

HPV vaccination is highly effective in the prevention of vaccine-type genital warts and cancers among MSM [11,12]. A study also showed that people who are already infected with one or more HPV types can still receive protection from other HPV types with the vaccines [13]. In addition, HPV vaccination could also prevent the development or recurrence of subsequent HPV-related diseases (eg, anal intraepithelial neoplasia) among people with a history of HPV infection [14-16]. Modeling work found that HPV vaccination for MSM could have a substantial impact on HPV-related disease incidence in this population and is cost-effective [17]. As a result, the US Centers for Disease Control and Prevention (CDC) recommends MSM who are in the age range of 45 years or below to receive HPV vaccination [13,18]. Free national HPV vaccination programs for MSM aged up to 45 years attending sexual health and HIV clinics rolled out in England and Scotland in 2018 and 2017, respectively [19,20]. The Victorian Government of Australia funded a free time-limited HPV vaccination catch-up program for MSM aged up to 26 years in 2017 [21]. The recommended course for MSM is 3 doses administered within 1 year; however, a 24-month period is clinically acceptable (second and third doses administered at least one month and 3 months after the first and second doses, respectively) [22]. However, in Hong Kong, no free or subsidized HPV vaccination program targeted MSM or other male populations. Two types of HPV vaccines (4-valent and 9-valent) are provided to males aged 9 years or more at the cost of HK $2500 to HK $4000 (US $321 to US $514) for a 3-dose course by private physicians in Hong Kong [23].

The uptake of free HPV vaccination provided by national programs was 37.6% among MSM in the United States [24], 42.6% in Australia [21], 49.1% in England [25], and 63.7% in Scotland [20]. HPV vaccination uptake among the male population is very low in Hong Kong [26], and previous studies have reported zero uptake among MSM [27,28]. About 30% of MSM in Hong Kong intended to take up HPV vaccination at the market rate [27,28]. Our literature search identified 4 studies that promoted HPV vaccination among MSM [29-32]. In a pilot intervention study conducted in the United States, young MSM were recruited via a popular web-based dating app and linked to a mobile health (mHealth) tool providing information and fostering access to HPV vaccination [31]. Of the 42 MSM who engaged with the mHealth tool, 11 (26%) received HPV vaccination [31]. Another web-based intervention was developed to provide information related to HPV prevalence among MSM in the United States, the effectiveness of the HPV vaccination, how to address potential barriers of receiving HPV vaccination, vaccine costs, and health insurance. A pilot randomized controlled trial (RCT) showed that such intervention was effective in increasing self-reported HPV vaccination initiation (receipt of 1 or more doses) compared with providing simple information about the vaccines (34/76, 45% vs 19/74, 26%; P=.02). However, the intervention did not increase HPV vaccination completion (receipt of 3 required doses: 8/76, 11% vs 2/74, 3%; P=.07). An ongoing RCT is comparing the efficacy of the same web-based intervention plus 2 different types of reminders (interactive vs noninteractive) versus the provision of simple HPV vaccination information among MSM in the United States [30]. In addition, an RCT tested the efficacy of texting-based intervention in the promotion of HPV vaccination among young MSM in the United States [32]. Participants who were assigned to the intervention group received daily SMS text messages for the first 3 weeks related to HPV vaccination, whereas those in the control group received SMS text messages related to other sexual health topics [32]. The intervention was effective in promoting HPV vaccination initiation (19.4% vs 6.6%) [32]. To our knowledge, no HPV vaccination promotion targeting MSM was conducted outside the United States.

A meta-analysis showed that theory-based interventions were more effective than nontheory-based interventions [33]. Our formative studies showed that some constructs of the health belief model (HBM) were significantly associated with the willingness to take up HPV vaccination at the market rate among

local MSM [27,28]. These constructs included perceived susceptibility (risk of HPV, genital warts, and penile or anal cancer), perceived severity (chance of HPV infection causing genital warts and penile or anal cancers), perceived benefit of HPV vaccination (efficacy in preventing genital warts and penile or anal cancers), perceived barriers (high cost, potential side effects, and embarrassment), and cue to action (recommended by doctor or peers to take up the vaccines). Such findings were considered while developing the web-based interventions used in this study. Video was produced to promote HPV vaccination, as the audio-visual approach has been used effectively in health promotion programs. The authors also conducted brief motivational interviewing (MI) over the phone to promote in one of the intervention groups. MI is a client-centered, nondirective, goal-oriented counseling technique, which helps clients explore and resolve any ambivalence that might have to change [34]. The US CDC lists MI among its best evidence intervention [35]. A systematic review indicated that MI was acceptable to MSM and feasible to deliver over the phone [36]. A telephonic delivery of MI is useful in changing HIV or sexually transmitted infection (STI)–related behaviors among MSM [37].

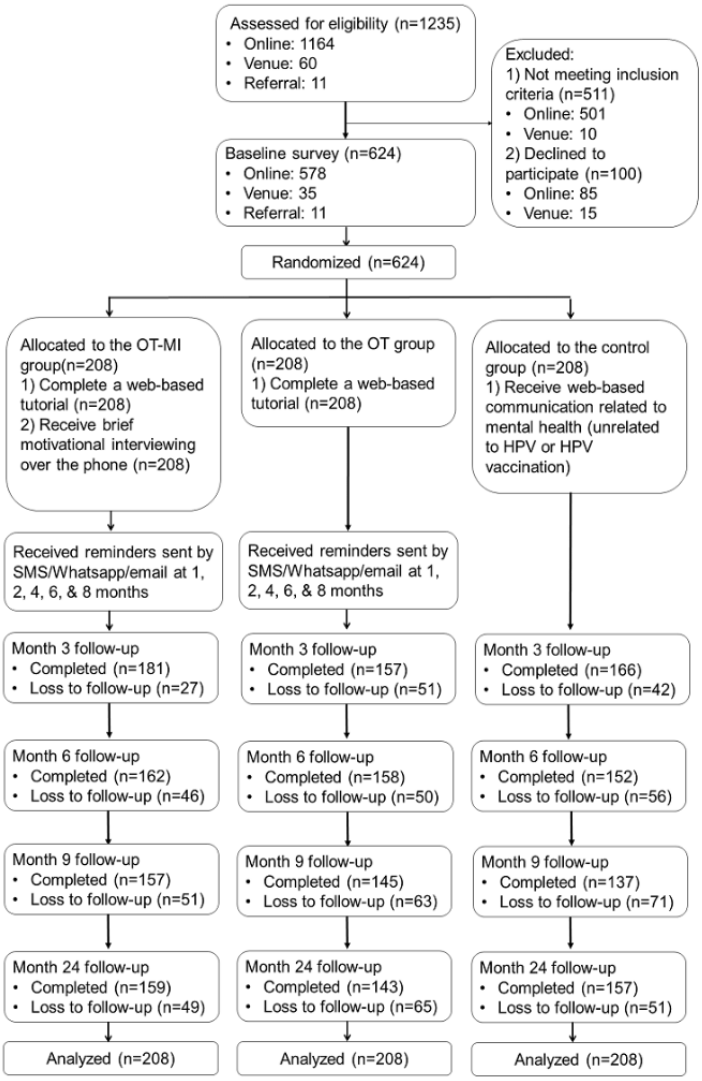
The primary objective of the RCT is to evaluate the relative efficacies of 2 theory- and web-based interventions with and without brief MI over the phone in increasing the completion of HPV vaccination (receipt of 3 required doses) within a 24-month follow-up period among unvaccinated Hong Kong Chinese MSM compared with the control group. The secondary objective is to evaluate the relative efficacies of the interventions in changing constructs related to the HBM over the follow-up period.

Methods

Study Design

A 3-arm parallel RCT was conducted between July 2017 and December 2019. This study was registered at ClinicalTrial.gov (number NCT03286907). The consort flowchart is shown in Figure 1.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the study.



Participants

Participants were (1) Hong Kong Chinese–speaking men aged between 18 and 45 years; (2) self-reported oral or anal intercourse with at least one man in the past 6 months; (3) having regular internet access; and (4) willing to complete telephonic follow-up evaluations. Those who had ever received HPV vaccination were excluded. We also excluded MSM aged over 45 years because HPV vaccination was not approved for use in this age group [18].

Recruitment Procedures

Participants were recruited through multiple sources. Recently, the Hong Kong government geographically located all venues frequently visited by local MSM, including 12 gay bars and 16 gay saunas [38]. Upon approval of the owners, trained and experienced fieldworkers approached prospective MSM participants in these venues at different time slots during weekdays and weekends. These fieldworkers briefed prospective participants about the study details and gave them an information sheet. The research team also conducted a web-based outreach by periodically posting information about the study as discussion topics on 2 gay websites with the highest traffic in Hong Kong. If prospective participants were interested in this study, they could contact the interviewers through private messaging or other means (eg, WhatsApp, telephone, and email). Recruitment was supplemented by peer referrals. Participants were guaranteed anonymity during the study and had the right to discontinue participation in the study at any time. Their refusal or withdrawal from the study would not affect their access to any future services. Verbal instead of written informed consent was obtained so as to maintaining anonymity, but the fieldworkers signed a form pledging that the participants had been fully informed about the study. Multiple forms of contact information were obtained to make an appointment to conduct a baseline telephonic interview. Upon appointment, trained telephone interviewers confirmed the eligibility and consent of the participants to participate in the study and conducted telephone interviews, which took about 10 min to complete. The follow-up surveys at 3, 6, and 9 months only recorded HPV vaccination uptake (number of doses, price, venue, date, and side effects). Participants who presented at 24 months were asked about HPV vaccination history and perceptions based on the HBM. Up to 5 calls were made at different timeslots during weekdays or weekends before considering a case as loss-to-follow-up. Upon completion of each of the 5 surveys, an HK $50 (US $6.3) supermarket or café coupon was mailed to participants as compensation for their time. Telephone numbers and addresses were cross-checked to avoid repetition. Ethics approval was obtained from the Survey and Behavioral Research Ethics Committee of the Chinese University of Hong Kong (ref no.: 13141651) and the Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee (ref no.: 2015.687-T).

Among 1235 prospective participants who were approached, none had received HPV vaccination. A total of 724 MSM were eligible for participation after screening, 100 refused to participate in the study for time and/or other logistical reasons, and 624 completed the baseline survey. The response rate was 86.2%. The number of participants who completed the follow-up surveys was 504 (504/624, 80.8%) at 3 months, 472 (472/624, 75.6%) at 6 months, 439 (439/624, 70.4%) at 9 months, and 459 (459/624, 73.6%) at 24 months, respectively.

Development of Intervention Materials

A panel comprising 3 MSM volunteers, 2 epidemiologists, 1 health psychologist, 1 health communication expert, 2 experienced nongovernmental organization (NGO) workers, and 1 video producer was formed. A literature review and a discussion group that involved 5 MSM were conducted to inform the design of the web-based health communication messages. The web-based tutorial video was produced by a professional team, reviewed by 3 other MSM, and finalized by the panel.

MI Training and Ongoing Supervision

Fieldworkers received 2 full days (14 hours in total) of basic MI training by an experienced trainer who is a member of the MI Network of Trainers. In addition, fieldworkers were trained to deliver the intervention using role-plays over the phone to increase their confidence. The role-plays and protocol-specific practice were conducted twice a week for 4 weeks and maintained by coaching sessions once every 2 weeks by the same trainer. Fieldworkers were not deployed in the fieldwork unless they had achieved a beginner level of competence. In addition, fieldworkers also received a 4-hour training workshop about background knowledge of local MSM, HPV-related diseases, and HPV vaccination. Meetings for problem solving and improvements were held 2 weeks after the commencement of fieldwork and monthly afterward. Fieldworkers sought support from the trainer via phone calls during the project period.

The Baseline Survey and Random Allocation Process

After completing the baseline telephone survey, the participants were randomized 1:1:1 to either the control group, the OT group, or the OT-MI group. Computer-generated random allocation codes were produced and sealed in opaque envelopes by a research staff member with no involvement in recruitment or baseline survey. One envelope was drawn and opened by the interviewers. They then informed the participant which group they were assigned to. Block randomization with a block size of 12 was used.

Intervention for the OT Group

A unique login identity and a link to access a web-based self-administered tutorial were sent to participants in the OT group. Next, the participants were requested to visit this website and performed the self-administered tutorial within 3 days. Up to 5 reminders (in the one-week period) were sent to those who did not log in within 3 days via email or social media channels.

The first part of the online tutorial (OT) was a 5-min web-based video. The carefully designed contents were based on theory-based factors associated with the acceptability of HPV vaccination among local MSM [27,28] and guided by the HBM [39]. In the video, a peer MSM discussed the high risk of having penile or anal cancers among MSM, severe consequences of genital warts, penile or anal cancers, promising efficacy, and long protection duration of HPV vaccination in the prevention

of genital warts and penile or anal cancers. Flashes of scary images of genital warts on penis/anal and penile and anal cancers were shown in the video to increase perceived severity. The peer MSM also emphasized that HPV vaccination is a worthy long-term investment for their health and demonstrated the procedures for receiving HPV vaccination in collaborative private clinics, which portrayed caring, guaranteed privacy, and a nonjudgmental environment. To ensure a complete exposure, the video was formatted in such a way that the participants could not fast-forward or skip any part of it.

The second part of the OT was a self-administered exercise. The participants answered some web-based multiple choice questions (eg, knowledge of HPV and HPV vaccination) and performed a short exercise to modify related cognitions (eg, "Please list three potential benefits you will gain after taking up HPV vaccination"). Model answers were displayed for incorrect entries, and the participants were requested to re-enter the correct items to complete the exercise. The entire tutorial takes about 20 min to complete.

Intervention for the OT-MI Group

On top of the same OT received by the OT group, participants in the OT-MI group received MI via telephone. The interviewer made an appointment with the participants to perform a 15-min MI session over the phone after completing the baseline survey.

Reminders for the OT and the OT-MI Groups

Reminders were sent via WeChat, WhatsApp, or emails to the participants of OT and OT-MI groups at 1, 2, 4, 6, and 8 months.

Intervention for the Control Group

Participants in the control group received a link to access web-based health communication messages about the prevalence of some common mental health problems among MSM and the introduction of stress reduction exercises, which was totally unrelated to HPV or HPV vaccination. They did not receive any reminders.

HPV Vaccination Uptake

If desired, then the interviewers would facilitate them to make an appointment to take up HPV vaccination in a collaborative private clinic at the market rate (HK $3800 [US $490] for 3 doses). No subsidy was provided. The authors did not limit participants' choice to take up HPV vaccination at other places.

Measures

Background Characteristics and Potential Confounders

Information collected included sociodemographic data (age, current marital status, educational level, current employment status, personal monthly income), sexual orientation, utilization of HIV or STI prevention services, history of HIV and other STIs, lifestyles (smoking and drinking), and sexual behaviors in the past 6 months. Queried sexual behaviors included anal intercourse with regular and nonregular sex partners, condomless anal intercourse (CAI) with men, multiple male sex partnerships, use of sexual potency drugs, and engagement in sexualized drug use. Regular male sex partners (RP) were defined as lovers and/or stable boyfriends, whereas nonregular partners (NRP) were defined as causal sex partners and/or male sex workers. In this study, sexualized drug use was defined as the use of any of the following psychoactive substances before or during sexual intercourse, including ketamine, methamphetamine, cocaine, cannabis, ecstasy, Dormicum/Halcion/Erimin 5/nonprescription hypnotic drugs, heroin, cough suppressant (not for curing cough), amyl nitrite (popper), GHB/GBL (γ-hydroxybutyrate), 5-methoxy-N, N-diisopropyltryptamine (Foxy), and mephedrone [40,41].

Primary Outcome

The primary outcome of the study was the validated completion of HPV vaccination within a 24-month follow-up period. To validate HPV vaccination uptake, the participants were requested to send the research team an image of their receipt hiding personal identification after they had received each dose of HPV vaccines. Supplemental information, including the location and cost of HPV vaccination uptake, was also collected.

Secondary Outcomes

Six scales assessed perceptions based on the HBM at baseline and 24 months. They were (1) the 4-item Perceived Susceptibility Scale; (2) the 2-item Perceived Severity Scale; (3) the 5-item Perceived Benefit Scale; (4) the 6-item Perceived Barrier Scale; (5) the 2-item Cue to Action Scale; and (6) the 3-item perceived self-efficacy scale. Cronbach α values for these 6 scales were acceptable (.72-.79). Single factors were identified for these scales in exploratory factor analysis, which explained 52.0%-71.0% of the total variance.

Postvaccination Experiences

Vaccinated participants were asked whether they had side effects of the vaccines (eg, pain, headache, fatigue, fever, nausea) and the severity of those side effects.

Process Evaluation

At the end of the OT, participants were requested to answer 3 simple questions about some details of the video to verify their exposure. A similar approach was used by some other published studies [37]. The computer program recorded the starting and ending times of the OT. The fieldworkers recorded the starting and ending times of the MI for verification.

The process evaluation of health promotion was conducted at the third month. Participants in OT and OT-MI groups were asked (1) whether the content of the web-based video was clear; (2) whether the OT was attractive for you; and (3) whether the OT had increased their understanding of the benefit of HPV vaccination and willingness to take up HPV vaccination. Participants in the OT-MI group were asked some additional questions about their satisfaction with the MI session.

At 24 months, the participants were asked whether they had been exposed to any health communication messages promoting HPV vaccination during the project period. Those with such exposure were asked about some details, such as time, sources, contents, and the framing of such messages (ie, whether they supported or were against HPV vaccination).

Sample Size Planning

Even without any intervention, about 30% intended to take up the HPV vaccine [27,28]. The meta-analysis showed that

43%-62% of those with a behavioral intention would translate it into action. In this study, the authors assumed that 12% of the participants in the OT group would complete HPV vaccination within the 24-month follow-up period. The authors expected that 2% of the participants in the control group would do so in the absence of any intervention. A sample size of 144 per group would allow us to detect a 10% (14/144) difference in the HPV vaccination uptake between the OT and control groups, and a 15% (22/144) difference between the OT-MI and OT groups (α=.0125; considering multiple comparisons, power of .8). With an expected loss-to-follow-up rate of 30% at 24 months, a sample size of approximately 208 per group is required (total n=624).

Statistical Analysis

Intention-to-treat (ITT) analysis was used for the outcome analyses. Missing data were handled by the imputation strategy of last-observation-carried-forward, a standard method in the ITT analysis [42]. The chi-square test (for categorical variables) and one-way ANOVA (for continuous variables) were used to inspect the between-group balances of baseline characteristics. Logistic regression models (for primary outcome) and multivariable linear regression models (for secondary outcomes) were used to test the between-group differences in primary and secondary outcomes, after controlling for any background variables, showed P<.20 in between-group comparisons. Adjusted odds ratios (AORs) and adjusted standardized coefficients (β) were also obtained.

Baron and Kenny's method was used to test the hypothesis that changes in the studied HBM constructs, which were used to design intervention that would mediate the between-group difference in the primary outcome [43]. Variables were constructed to assess changes in the HBM constructs (24 months-baseline). The mediation hypothesis was tested by first inspecting the association between changes in HBM constructs and randomization status. Logistic regression models were then fit using the validated uptake of 3 doses of HPV vaccination as the dependent variable and adjusted for significant baseline characteristics measured at baseline. Model 1 included randomization status as an independent variable, Model 2 included the changes in HBM constructs as the independent variables, and Model 3 included variables of both Models 1 and 2. If the association between randomization status and HPV vaccination uptake became statistically nonsignificant after adjusting for the changes in HBM constructs, the changes in HBM constructs were observed to fully mediate the association between randomization status and the dependent variable. Partial mediation occurred when the strength of association between randomization status and the dependent variable was reduced when remaining statistically significant changes in HBM constructs were controlled. SPSS version 16.0 was used, and P values <.05 were considered statistically significant.

Results

Descriptive Statistics

At baseline, the majority of the participants were single (510/624, 81.7%), full-time employed (492/624, 78.8%), had attended college or above (529/624, 84.8%), and identified themselves as gay (557/624, 89.3%). About one-third were aged between 18 and 26 years (202/624, 32.4%), and 50.2% (313/624) reported a monthly income of HK $20,000 (US $2580) or above. Regarding service utilization, 55.4% (346/624) and 48.1% (300/624) used HIV testing and other HIV or STI preventive services in the past 6 months. In the past 6 months, 80.4% (502/624) and 51.4% (321/324) had anal intercourse with RP and NRP, 35.3% (220/624) reported CAI with any men, 6.6% (41/624) used sexual potency drugs before or during anal intercourse, and 6.3% (39/624) reported sexualized drug use. The level of knowledge and perceptions related to HPV or HPV vaccination are shown in Table 1. Apart from anal intercourse with RP (P=.03) and the score of the perceived barrier scale (P=.03), no significant between-group difference was found (P=.24 to P=.96). Therefore, anal intercourse with RP and the score of the Perceived Barrier Scale were controlled in the subsequent analysis of primary and secondary outcomes.

Table 1. Characteristics of participants at baseline.

Characteristics	Control group (n=208)	OT[a] group (n=208)	OT-MI[b] group (n=208)	*P* value (comparing difference among control/OT/OT-MI groups)
Sociodemographic data, n (%)				
Age group (year)				.57
18-26	65 (31.3)	73 (35.1)	64 (30.8)	
27-36	101 (48.6)	91 (43.8)	91 (43.8)	
37-45	42 (20.2)	44 (21.2)	53 (25.5)	
Current marital status				.24
Currently single	171 (82.2)	162 (77.9)	177 (85.1)	
Cohabited/married with a man	35 (16.8)	45 (40.5)	31 (14.9)	
Cohabited/married with a woman	2 (1.0)	1 (0.5)	0 (0.0)	
Educational level				.71
Secondary or below	29 (13.9)	31 (14.9)	35 (16.8)	
University or above	179 (86.1)	177 (85.1)	173 (83.2)	
Current employment status				.82
Full-time	166 (79.8)	165 (79.3)	161 (77.4)	
Part-time/unemployed/retired/students	42 (20.2)	43 (20.7)	47 (22.6)	
Personal monthly income (HK $; US $)				.38
<HK $10,000 (US $1290)	32 (15.4)	36 (17.3)	33 (15.9)	
HK $10,000-$19,999 (US $1290-$2580)	74 (35.6)	71 (34.1)	62 (29.8)	
HK $20,000-$39,999 (US $2580-$5161)	79 (38.0)	67 (32.2)	77 (37.0)	
≥HK $40,000 (US $5161)	21 (10.1)	34 (16.3)	35 (16.8)	
Refuse to disclose	2 (1.0)	0 (0.0)	1 (0.5)	
Sexual orientation				.31
Gay	179 (86.1)	189 (90.9)	189 (90.9)	
Bisexual	28 (13.5)	19 (9.1)	19 (9.1)	
Heterosexual	1 (0.5)	0 (0.0)	0 (0.0)	
HIV or STI[c]-related service use in the past 6 months, n (%)				
HIV testing				.55
No	93 (44.7)	98 (47.1)	87 (41.8)	
Yes	115 (55.3)	110 (52.9)	121 (58.2)	
Other HIV or STI preventive services[d]				.55
No	102 (49.0)	109 (52.4)	113 (54.3)	
Yes	106 (51.0)	99 (47.6)	95 (45.7)	
Sexual behaviors in the past six months , n (%)				
Had anal intercourse with RP[e]				.03
No	31 (14.9)	39 (18.8)	52 (25.0)	
Yes	177 (85.1)	169 (81.3)	156 (75.0)	
Had anal intercourse with NRP[f]				.93
No	103 (49.5)	101 (48.6)	99 (47.6)	

Characteristics	Control group (n=208)	OT[a] group (n=208)	OT-MI[b] group (n=208)	*P* value (comparing difference among control/OT/OT-MI groups)
Yes	105 (50.5)	107 (51.4)	109 (52.4)	
CAI[g] with men				.74
No	133 (63.9)	139 (66.8)	132 (63.5)	
Yes	75 (36.1)	69 (33.2)	76 (36.5)	
Multiple male sex partnerships				.88
No	90 (43.3)	92 (44.2)	95 (45.7)	
Yes	118 (56.7)	116 (55.8)	113 (54.3)	
Sexualized drug use				.56
No	193 (92.8)	198 (95.2)	194 (93.3)	
Yes	15 (7.2)	10 (4.8)	14 (6.7)	
Use of sexual potency drugs				.66
No	193 (92.8)	197 (94.7)	193 (92.8)	
Yes	15 (7.2)	11 (5.3)	15 (7.2)	
History of HIV or STI, n (%)				
Self-reported HIV serostatus				.96
Negative	183 (88.0)	185 (88.9)	184 (88.5)	
Positive	11 (5.3)	7 (3.4)	8 (3.8)	
Refuse to disclose	3 (1.4)	4 (1.9)	5 (2.4)	
Had never tested for HIV antibody	11 (5.3)	12 (5.8)	11 (5.3)	
History of other STIs				.61
No	170 (81.7)	170 (81.7)	163 (26.1)	
Yes	38 (18.3)	38 (18.3)	45 (21.6)	
Lifestyles, n (%)				
Current smokers				.88
No	165 (79.3)	162 (77.9)	166 (79.8)	
Yes	43 (20.7)	46 (22.1)	42 (20.2)	
Drinking in the past year				.32
No	39 (18.8)	28 (13.5)	36 (17.3)	
Yes	169 (81.2)	180 (86.5)	172 (82.7)	
Knowledge related to HPV[h] or HPV vaccination, n (%)				
Both males and females could be affected by HPV				.06
Yes[i]	156 (75.0)	174 (83.7)	163 (78.4)	
No	9 (4.3)	11 (5.3)	15 (7.2)	
Do not know	43 (20.7)	23 (11.1)	30 (14.4)	
HPV infection could cause STI				.63
Yes[i]	129 (62.0)	144 (69.2)	135 (64.9)	
No	20 (9.6)	18 (8.7)	19 (9.1)	
Do not know	59 (28.4)	46 (22.1)	54 (26.0)	
HPV infection could cause cancers among males				.07
Yes[i]	89 (42.8)	114 (54.8)	99 (47.6)	

Characteristics	Control group (n=208)	OT[a] group (n=208)	OT-MI[b] group (n=208)	*P* value (comparing difference among control/OT/OT-MI groups)
No	34 (16.3)	37 (17.8)	37 (17.8)	
Do not know	85 (40.9)	57 (27.4)	72 (34.6)	
HPV could be totally cured by available treatment				.16
Yes	31 (14.9)	28 (13.5)	35 (16.8)	
No[i]	110 (52.9)	132 (63.5)	112 (53.8)	
Do not know	67 (32.2)	48 (23.1)	61 (29.1)	
Availability of effective HPV vaccination for males in Hong Kong				.03
Yes[i]	113 (54.3)	131 (63.0)	110 (52.9)	
No	18 (8.7)	28 (13.5)	26 (12.5)	
Do not know	77 (37.0)	49 (23.6)	72 (34.6)	
Number of shots required to prevent HPV infection in males				.31
3	56 (26.9)	69 (33.2)	68 (32.7)	
Other answers/Do not know	152 (73.1)	139 (66.8)	140 (67.3)	
Number of correct responses				.26
0	33 (15.9)	19 (9.1)	27 (13.0)	
1-2	45 (21.6)	40 (19.2)	46 (22.1)	
3-4	100 (48.1)	104 (50.0)	95 (47.9)	
5-6	30 (14.4)	45 (21.6)	40 (19.2)	
Perceptions related to HPV or HPV vaccination based on the HBM[j]				
Perceived susceptibility to HPV (high/very high)				
Perceived risk of contracting HPV in lifetime, n (%)	42 (20.2)	52 (25.0)	49 (23.6)	.49
Perceived risk of contracting genital warts in lifetime, n (%)	39 (18.8)	51 (24.5)	47 (22.6)	.35
Perceived risk of having penile or anal cancers in lifetime, n (%)	20 (9.6)	25 (12.0)	21 (10.1)	.70
Perceived HPV infection rate among MSM[k] in Hong Kong, n (%)	63 (30.3)	53 (25.5)	55 (26.4)	.51
Perceived susceptibility scale, mean (SD)[i]	10.8 (3.0)	10.9 (3.3)	10.4 (3.5)	.38
Perceived severity of HPV-related diseases (high/very high)				
Harms of genital warts on physical health, n (%)	119 (57.2)	119 (57.2)	128 (61.5)	.59
Harms of penile or anal cancers on physical health, n (%)	135 (64.9)	153 (73.6)	137 (65.9)	.12
Perceived severity scale, mean (SD)[l]	7.6 (1.9)	7.7 (1.7)	7.5 (1.8)	.59
Perceived benefit of HPV vaccination (agree/strongly agree)				
HPV vaccination is highly effective in preventing HPV infection, n (%)	143 (68.8)	166 (79.8)	147 (70.7)	.03

Characteristics	Control group (n=208)	OT[a] group (n=208)	OT-MI[b] group (n=208)	*P* value (comparing difference among control/OT/OT-MI groups)
HPV vaccination is highly effective in preventing genital warts, n (%)	142 (68.3)	151 (72.6)	138 (66.3)	.37
HPV vaccination is highly effective in preventing penile/anal cancers, n (%)	129 (62.0)	138 (66.3)	123 (59.1)	.31
HPV vaccination can protect you for a long time, n (%)	104 (50.0)	113 (54.3)	105 (50.5)	.63
You will feel at ease after receiving HPV vaccination, n (%)	142 (68.3)	127 (61.1)	139 (66.8)	.26
Perceived benefit scale, mean (SD)[m]	18.9 (3.1)	18.9 (2.9)	18.6 (2.9)	.58
Perceived barriers of receiving HPV vaccination (agree/strongly agree)				
It is not worthy spending HK $2000-$3000 (US $257.97-$386.96) to receive HPV vaccination, n (%)	64 (30.8)	44 (21.1)	38 (18.3)	.01
The procedures to receive HPV vaccination is troublesome, n (%)	33 (15.9)	30 (13.3)	17 (8.2)	.05
You would have severe side effects after receiving HPV vaccination, n (%)	20 (9.6)	15 (7.2)	17 (8.2)	.67
You feel embarrassed to receive HPV vaccination, n (%)	25 (12.0)	20 (9.6)	18 (8.7)	.50
Others would think you are having high-risk behaviors if you receive HPV vaccination, n (%)	33 (15.9)	25 (12.0)	25 (12.0)	.41
You would be stigmatized by service providers when you receive HPV vaccination, n (%)	18 (8.7)	15 (7.2)	13 (6.3)	.64
Perceived barrier scale, mean (SD)[n]	13.5 (5.2)	12.6 (4.0)	12.4 (4.2)	.03
Perceived cue to action related to HPV vaccination (agree/strongly agree)				
Medical professionals would suggest you to receive HPV vaccination, n (%)	3 (1.4)	6 (2.9)	7 (3.4)	.43
MSM peers would suggest you to receive HPV vaccination, n (%)	8 (3.8)	15 (7.2)	5 (2.4)	.06
Cue to action scale, mean (SD)[o]	2.7 (1.3)	2.9 (1.5)	2.8 (1.4)	.56
Perceived self-efficacy related to HPV vaccination (agree/strongly agree)				
Whether to receive HPV vaccination is completely under your control, n (%)	172 (82.7)	169 (81.3)	183 (88.0)	.14
You are confident to receive HPV vaccination in the next year if you want, n (%)	142 (68.3)	139 (66.8)	147 (70.7)	.70

Characteristics	Control group (n=208)	OT[a] group (n=208)	OT-MI[b] group (n=208)	*P* value (comparing difference among control/OT/OT-MI groups)
Receiving HPV vaccination in the next year is easy for you if you want, n (%)	154 (74.0)	145 (69.7)	153 (73.6)	.56
Perceived self-efficacy scale, mean (SD)[p]	12.6 (2.4)	12.4 (2.6)	12.6 (2.4)	.61

[a]OT: online tutorial.

[b]MI: motivational interviewing.

[c]STI: sexually transmitted infection.

[d]Including receiving condoms, peer education, leaflets of HIV-related information, and seminars or workshops related to HIV.

[e]RP: regular male sex partners.

[f]NRP: nonregular partners.

[g]CAI: condomless anal intercourse.

[h]HPV: human papillomavirus.

[i]Perceived susceptibility scale, 3 items, Cronbach α=.79, one factor was identified by exploratory factor analysis, which explained 61.5% of the total variance.

[j]HBM: health belief model.

[k]MSM: men who have sex with men.

[l]Perceived severity scale, 2 items, Cronbach α=.64.

[m]Perceived benefit scale, 5 items, Cronbach α=.75, one factor was identified by exploratory factor analysis, explaining 52.0% of the total variance.

[n]Perceived barrier scale, 6 items, Cronbach α=.86, one factor was identified by exploratory factor analysis, explaining 71.0% of the total variance.

[o]Cue to action scale, 2 items, Cronbach α=.62.

[p]Perceived self-efficacy scale: 3 items, Cronbach α=.78, one factor was identified by exploratory factor analysis, explaining 69.2% of total variance.

The loss-to-follow-up rates in the control, OT, and OT-MI groups at 24 months were 24.5% (51/208), 31.3% (65/208), and 23.6% (49/208), respectively. Significant differences in the current marital status, educational level, anal intercourse with NRP, CAI with men, multiple male sex partnerships, sexualized drug use, and the use of sexual potency drugs were found in at least one group while comparing those who were followed up and those who were lost to follow-up at 24 months (Multimedia Appendix 1).

Between-Group Difference in Primary Outcome

All 75 participants who self-reported having had completed HPV vaccination at 24 months were able to provide receipts for verification. Participants in the OT-MI group reported significantly higher HPV vaccination completion rates at 24 months than that by the control group (36/208, 17.3% vs 15/208, 7.2%; AOR 1.57, 95% CI 1.14-2.17; *P*=.006). However, the difference in HPV vaccination completion between the OT and the control groups (24/208, 11.5% vs 15/208, 7.2%; AOR 1.61, 95% CI 0.82-3.18; *P*=.17), or between OT-MI and OT groups (36/208, 17.3% vs 24/208, 11.5%, AOR 1.55, 95% CI 0.89-2.72; *P*=.13), was not statistically significant.

The location for receiving HPV vaccination included the collaborative private clinic (54/75, 72%), other private clinics in Hong Kong (19/75, 25%), university health care center (1/75, 1%), and clinics in Australia (1/75, 1%). Except for one participant who received free HPV vaccination in Australia, other participants self-paid HK $1600 to HK $9000 (US $206-US $1161; median: HK $3800 or US $490) to receive the vaccination. The majority of the vaccinated participants completed the entire course within 12 months (55/75; 73%), whereas 20 (20/75, 27%) completed it within 24 months.

Between-Group and Within-Group Differences in Secondary Outcomes

Compared with the participants in the control group, participants in the OT-MI (adjusted β: –.21; *P*<.001) and OT (adjusted β: –.10; *P*=.02) groups reported lower perceived barriers to taking up HPV vaccination at 24 months. The comparison of 24-month and baseline data revealed statistical increase in perceived severity (in all 3 groups, *P*<.001) and perceived barriers (in the control group and the OT group, *P*<.001 and *P*=.001, respectively). A statistically significant decrease in perceived benefit was found in the control group (*P*=.04), whereas a decrease in perceived self-efficacy was found in both the control (*P*=.05) and the OT-MI groups (*P*=.002; Table 2 and Multimedia Appendix 1).

Table 2. Between-group comparisons of secondary outcomes.

Secondary outcomes (perceptions based on the HBM[a])	OT[b] group vs control group		OT-MI[c] group vs control group		OT-MI group vs OT group	
	Adjusted β[d]	*P* value	Adjusted β	*P* value	Adjusted β	*P* value
Perceived susceptibility scale						
Baseline	.01	.91	–.07	.19	–.06	.22
24 months	–.01	.88	.05	.26	.06	.23
24 months-Baseline	–.01	.79	.12	.02	.11	.02
Perceived severity scale						
Baseline	.04	.37	.01	.82	–.04	.38
24 months	.01	.80	.03	.62	.02	.68
24 months-Baseline	–.03	.52	.02	.75	.06	.24
Perceived benefit scale						
Baseline	–.02	.70	–.06	.20	–.05	.30
24 months	–.001	.98	.03	.53	.02	.69
24 months-Baseline	.02	.75	.09	.09	.07	.19
Perceived barrier scale						
Baseline	–.10	.05	–.12	.02	–.03	.60
24 months	–.11	.01	–.21	<.001	–.10	.02
24 months-Baseline	–.12	.01	–.21	<.001	–.11	.02
Cue to action scale						
Baseline	.06	.26	.02	.68	–.04	.48
24 months	–.004	.93	.01	.78	.01	.87
24 months-Baseline	–.05	.34	–.003	.95	.03	.49
Perceived self-efficacy scale						
Baseline	–.08	.10	–.05	.29	.03	.55
24 months	–.06	.22	–.09	.08	–.04	.45
24 months-Baseline	.02	.66	–.03	.53	–.06	.23

[a]HBM: health belief model.

[b]OT: online tutorial.

[c]MI: motivational interviewing.

[d]Standardized coefficients adjusted for anal intercourse with regular male sex partners and score of Perceived Barrier Scale measured at baseline.

Testing the Mediation Hypotheses

Adjusted for potential confounders assessed at baseline, the within-group changes in the scores of the Perceived Susceptibility Scale and Perceived Barrier Scale were significantly associated with the intervention status (the OT-MI group vs the control group, *P*=.02 and *P*<.001, respectively) and HPV vaccination completion (*P*<.001, Table 3).

The association between the intervention status (OT-MI vs control) and HPV vaccination completion became statistically nonsignificant (AOR 1.42, 95% CI 0.99-2.03; *P*=.06) after controlling for the change in perceived susceptibility, whereas the change in perceived susceptibility remained strongly associated with the dependent variable in the model (AOR 1.23, 95% CI 1.12-1.35; *P*<.001). The results suggested that changes in perceived susceptibility fully mediated the intervention effect. The association between intervention status and HPV vaccination completion was weakened (from *P*<.001 to *P*=.05) when changes in perceived barriers were controlled in the model, with the change in perceived barrier remaining statistically significant in the model (*P*<.001). A partial mediation effect was observed (Table 3).

Table 3. Test for the independent effect of changes in Health Belief Model scale scores (24 months vs baseline) on the association between intervention status (online tutorial- motivational interviewing group vs control group) and human papillomavirus vaccination completion during the follow-up period (n=416).

Model and variables	Baseline	SE	AOR[a,b] (95% CI)	*P* value
1				
Intervention status (OT[c]-MI[d] group vs control group)	0.47	0.17	1.61 (1.14-2.26)	.006
2A				
Change in score of the perceived susceptibility scale	0.22	0.05	1.25 (1.14-1.38)	<.001
3A				
Intervention status (OT-MI group vs control group)	0.35	0.18	1.42 (0.99-2.03)	.06
Change in score of the perceived susceptibility scale	0.21	0.05	1.23 (1.12-1.35)	<.001
2B				
Change in score of the perceived barrier scale	–0.18	0.04	0.84 (0.77-0.91)	<.001
3B				
Intervention status (OT-MI group vs control group)	0.37	0.18	1.44 (1.01-2.06)	.05
Change in score of the perceived barrier scale	–0.16	0.04	0.85 (0.79-0.92)	<.001

[a]AOR: adjusted odds ratios

[b]Odds ratios adjusted for potential confounders measured at baseline (sociodemographic data, HIV and sexually transmitted disease [STI]–related service utilization, history of HIV and STI, sexual behaviors, lifestyles, and knowledge related to human papillomavirus [HPV]/HPV vaccination).

[c]OT: online tutorial.

[d]MI: motivational interviewing.

Postvaccination Experiences

The majority of the vaccinated participants reported no side effects (65/75, 87%). The reported side effects included pain at the injection site (4/75, 5%), fever (3/75, 4%), numbness in arms (2/75, 3%), nausea (2/75, 3%), fatigue (1/75, 1%), and headache (1/75, 1%). Most of these side effects were mild (8/10; 80%).

Process Evaluation of Web-based Intervention Promoting HPV Vaccination

As recorded by the computer program, all participants in the OT and the OT-MI groups completed the OT. The time spent on the tutorial ranged from 15 to 28 min. The duration of MI ranged from 11 to 19 min.

Among those in the OT and OT-MI groups who participated in the process evaluation at 3 months, 84.1% (281/334) and 46.1% (154/334) believed that the content of the web-based health promotion video was clear and attractive. Moreover, 77.2% (258/334) and 47.0% (157/334) indicated that the OT has increased their understanding of HPV vaccination and willingness to take up HPV vaccination. Among those in the OT-MI group, 86.8% (159/182) were satisfied with the MI session, and 60.9% (111/182) believed that the MI session reduced their barriers toward getting HPV vaccination.

Among participants who completed the survey at 24 months 6.5% (30/459) had been exposed to other health communication messages that supported HPV vaccination during the project period, whereas 3.1% (14/459) had been exposed to other health communication messages that were against HPV vaccination during the same period. The most common resource of these health communication messages was pamphlets, followed by communication with peers and web-based communication.

Discussion

Principal Findings

Compared with the control group, the OT plus MI over the phone brought a significant increase in HPV vaccination completion among MSM over a 24-month follow-up period (36/208, 17.3% vs 15/208, 7.2%; *P*=.006). However, the hypothesis that using OT alone would show superior results as compared with the control group (24/208, 11.5% vs 15/208, 7.2%; *P*=.17) and adding MI to the OT would be more efficacious (36/208, 17.3% vs 24/208, 11.5%; *P*=.13) were not supported by the results.

The net increase in HPV vaccination uptake observed in this study (OT-MI group vs control group, 10.1%) was slightly higher than that of the other web-based interventions targeting MSM in the United States (8%) [29]. Unlike in the United States, where the cost of HPV vaccination can be fully covered by multiple sources of private and public financing [44], users self-paid for HPV vaccination at the market price without any subsidy in this study. Given the high HPV-related disease burden

among Chinese MSM, instead of waiting for the free HPV vaccination program to become available, there is an urgent need for promoting self-paid HPV vaccination among MSM in Hong Kong. For the first time, health workers in Hong Kong were provided with an evidence-based intervention for promoting HPV vaccination among MSM. The authors' intervention had some strengths. First, no additional funding is required to pay for HPV vaccination in this intervention. Second, the OT required minimum resources for maintenance. With simple training, health workers can easily perform brief MI. Therefore, governmental and NGOs can integrate this intervention into their existing HIV or STI prevention services and scale it up with little extra cost. The research team successfully embedded the intervention into HIV testing and counseling services run by a local NGO [45]. MSM using the service completes the OT on a tablet while they are waiting for their testing results. The NGO also enhanced posttest counseling to include brief MI promoting HPV vaccination for the users.

Given that the HPV vaccination uptake in the control group was higher than expected, and the difference between OT and control groups was smaller than expected, we believe that the nonsignificant findings may be attributed to an inadequate sample size and a limited statistical power. The *P* values for the comparison between the OT and control groups, and between the OT-MI and OT groups were close to .10. Future studies with larger sample sizes should be conducted. Therefore, the authors recommend OT together with MI for HPV vaccination promotion among MSM in Hong Kong until further evidence is generated.

The intervention also caused between-group differences in perceptions related to the HBM that were used to develop the intervention. The mediation analysis results explained some plausible mechanisms that might cause the observed behavioral change. The behavioral change may be caused by an increase in perceived susceptibility to HPV or HPV-related diseases. This finding was consistent with studies that demonstrated a strong association between risk perception and behavioral change in different health topics [46,47]. In the absence of intervention, perceived barriers to taking up HPV vaccination increased over time in the control group. Receiving OTs and brief MI over the phone slowed down the increase in perceived barriers. The between-group (OT-MI vs control) difference in changes of perceived barriers partially mediated the intervention effects. This study extended the applicability of HBM.

This study had the strength of its RCT design, long follow-up duration, well-validated primary outcome, and was based on theory and supported by the results of formative studies. The intervention was well received based on the positive process evaluation results. However, this study has some limitations. First, the intervention was limited to MSM who had internet access, but the majority of MSM in Hong Kong is expected to have access to the internet as the smartphone ownership is above 96% [48]. It is justifiable that the penetration of smartphones has been increasing sharply in many countries. Second, probability sampling was not feasible for this study in the absence of a sampling frame. Like most RCTs, the participants were recruited by convenient sampling, and selection bias might exist. The RCT design ensured a good internal validity. However, caution should be exercised while generalizing the results to other Chinese cities. Third, attrition bias might exist. The dropouts in OT-MI group had a higher prevalence of sexual risk behaviors than nondropouts, and they may have a higher motivation to take up HPV vaccination. The HPV uptake rate in the OT-MI group might be underestimated. Moreover, we did not collect information about MSM who refused to participate in the study. Furthermore, evaluating the MI sessions by audiotaping was the gold standard to assess fidelity. However, the authors were not able to do so, as the participants were MSM, and the studied questions covered sensitive topics such as their HIV and STI status and sexual behaviors.

Conclusions

The RCT findings showed that the theory-based OT, together with brief MI over the phone, was effective in increasing the completion of 3 required doses of HPV vaccination among MSM in Hong Kong, China. At present, local NGOs have integrated interventions in their HIV testing and counseling services for MSM. Local and international dissemination and implementation research are greatly warranted.

Acknowledgments

The research team would like to thank all the participants who participated in the study. Importantly, sincere thanks are extended to all those who had helped in the entire data collection period. The study was supported by the Health and Medical Research Fund, Food and Health Bureau, Hong Kong Special Administrative Region (Project Ref: #13141651).

Authors' Contributions

ZW and JL contributed equally as corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparing baseline characteristics between participants being followed up and those who were lost-to-follow-up at Month 24 and descriptive statistics of primary and secondary outcomes at baseline and Month 24 follow-up.
[DOCX File , 25 KB - jmir_v23i2e21465_app1.docx]

References

1. World Health Organziation. Human papillomavirus vaccines: WHO position paper, May 2017. Available at. URL: http://apps.who.int/iris/bitstream/handle/10665/255353/WER9219.pdf;jsessionid=0F6E736DC1E29E420436DE39CD4551DD?sequence=1 [accessed 2021-01-08]
2. Machalek DA, Poynten M, Jin F, Fairley CK, Farnsworth A, Garland SM, et al. Anal human papillomavirus infection and associated neoplastic lesions in men who have sex with men: a systematic review and meta-analysis. The Lancet Oncology 2012 May;13(5):487-500. [doi: 10.1016/s1470-2045(12)70080-3]
3. Chow EPF, Tucker JD, Wong FY, Nehl EJ, Wang Y, Zhuang X, et al. Disparities and risks of sexually transmissible infections among men who have sex with men in China: a meta-analysis and data synthesis. PLoS One 2014 Feb 24;9(2):e89959 [FREE Full text] [doi: 10.1371/journal.pone.0089959] [Medline: 24587152]
4. Liu F, Hang D, Deng Q, Liu M, Xi L, He Z, et al. Concurrence of oral and genital human papillomavirus infection in healthy men: a population-based cross-sectional study in rural China. Sci Rep 2015 Oct 27;5(1):15637 [FREE Full text] [doi: 10.1038/srep15637] [Medline: 26503510]
5. Hebnes JB, Olesen TB, Duun-Henriksen AK, Munk C, Norrild B, Kjaer SK. Prevalence of genital human papillomavirus among men in Europe: systematic review and meta-analysis. J Sex Med 2014 Nov;11(11):2630-2644. [doi: 10.1111/jsm.12652] [Medline: 25088239]
6. Zhang X, Yu J, Li M, Sun X, Han Q, Li M, et al. Prevalence and related risk behaviors of HIV, syphilis, and anal HPV infection among men who have sex with men from Beijing, China. AIDS Behav 2013 Mar 11;17(3):1129-1136. [doi: 10.1007/s10461-011-0085-x] [Medline: 22076229]
7. Jiang J, Cao N, Zhang J, Xia Q, Gong X, Xue H, et al. High prevalence of sexually transmitted diseases among men who have sex with men in Jiangsu Province, China. Sex Transm Dis 2006 Feb;33(2):118-123. [doi: 10.1097/01.olq.0000199763.14766.2b] [Medline: 16432484]
8. D Souza G, Wentz A, Wiley D, Shah N, Barrington F, Darragh TM, et al. Anal Cancer Screening in Men Who Have Sex With Men in the Multicenter AIDS Cohort Study. JAIDS Journal of Acquired Immune Deficiency Syndromes 2016;71(5):570-576. [doi: 10.1097/qai.0000000000000910]
9. Qin Q, Tang W, Ge L, Li D, Mahapatra T, Wang L, et al. Changing trend of HIV, Syphilis and Hepatitis C among Men Who Have Sex with Men in China. Sci Rep 2016 Aug 18;6(1):31081 [FREE Full text] [doi: 10.1038/srep31081] [Medline: 27535092]
10. Lin C, Lau JT, Ho K, Lau M, Tsui H, Lo K. Incidence of genital warts among the Hong Kong general adult population. BMC Infect Dis 2010 Sep 17;10(1):272 [FREE Full text] [doi: 10.1186/1471-2334-10-272] [Medline: 20849578]
11. Robinson CL, Advisory Committee on Immunization Practices (ACIP), ACIP Child/Adolescent Immunization Work Group. Advisory Committee on Immunization Practices Recommended Immunization Schedules for Persons Aged 0 Through 18 Years--United States, 2016. MMWR Morb Mortal Wkly Rep 2016 Feb 05;65(4):86-87 [FREE Full text] [doi: 10.15585/mmwr.mm6504a4] [Medline: 26845283]
12. De Vuyst H, Clifford GM, Nascimento MC, Madeleine MM, Franceschi S. Prevalence and type distribution of human papillomavirus in carcinoma and intraepithelial neoplasia of the vulva, vagina and anus: a meta-analysis. Int J Cancer 2009 Apr 01;124(7):1626-1636 [FREE Full text] [doi: 10.1002/ijc.24116] [Medline: 19115209]
13. HPV vaccine recommendations. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/vaccines/vpd/hpv/hcp/recommendations.html [accessed 2021-01-08]
14. Joura EA, Garland SM, Paavonen J, Ferris DG, Perez G, Ault KA, FUTURE III Study Group. Effect of the human papillomavirus (HPV) quadrivalent vaccine in a subgroup of women with cervical and vulvar disease: retrospective pooled analysis of trial data. BMJ 2012 Mar 27;344(mar27 3):e1401-e1401 [FREE Full text] [doi: 10.1136/bmj.e1401] [Medline: 22454089]
15. Swedish KA, Factor SH, Goldstone SE. Prevention of recurrent high-grade anal neoplasia with quadrivalent human papillomavirus vaccination of men who have sex with men: a nonconcurrent cohort study. Clin Infect Dis 2012 Apr 30;54(7):891-898. [doi: 10.1093/cid/cir1036] [Medline: 22291111]
16. Giuliano AR, Palefsky JM, Goldstone S, Moreira ED, Penny ME, Aranda C, et al. Efficacy of Quadrivalent HPV Vaccine against HPV Infection and Disease in Males. N Engl J Med 2011 Feb 03;364(5):401-411. [doi: 10.1056/nejmoa0909537]
17. Lin A, Ong KJ, Hobbelen P, King E, Mesher D, Edmunds WJ, et al. Impact and Cost-effectiveness of Selective Human Papillomavirus Vaccination of Men Who Have Sex With Men. Clin Infect Dis 2017 Mar 01;64(5):580-588 [FREE Full text] [doi: 10.1093/cid/ciw845] [Medline: 28011615]
18. Evidence to Recommendations for HPV Vaccination of Adults, Ages 27 through 45 years. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/vaccines/acip/recs/grade/HPV-adults-etr.html [accessed 2021-01-08]
19. HPV vaccination for men who have sex with men (MSM) programme. Public Health England. URL: https://www.gov.uk/government/collections/hpv-vaccination-for-men-who-have-sex-with-men-msm-programme [accessed 2021-01-08]
20. Pollock K, Wallace L, Wrigglesworth S, McMaster D, Steedman N. HPV vaccine uptake in men who have sex with men in Scotland. Vaccine 2019 Sep 03;37(37):5513-5514. [doi: 10.1016/j.vaccine.2018.11.081] [Medline: 30545714]

21. McGrath L, Fairley CK, Cleere EF, Bradshaw CS, Chen MY, Chow EPF. Human papillomavirus vaccine uptake among young gay and bisexual men who have sex with men with a time-limited targeted vaccination programme through sexual health clinics in Melbourne in 2017. Sex Transm Infect 2018 Nov 08;95(3):181-186. [doi: 10.1136/sextrans-2018-053619]
22. Public Health England. HPV vaccination pilot for men who have sex with men (MSM) 2016. Public Health England. URL: https://www.gov.uk/government/publications/hpv-vaccination-pilot-for-men-who-have-sex-with-men-msm [accessed 2021-01-08]
23. Consensus Statement on the use of Human Papillomavirus(HPV) Vaccine in prevention of cervical cancer. Centre for Health Protection. 2016. URL: https://www.chp.gov.hk/files/pdf/consensus_statement_on_the_use_of_hpv_vaccine_in_prevention_of_cervical_cancer.pdf [accessed 2021-01-08]
24. Loretan C, Chamberlain AT, Sanchez T, Zlotorzynska M, Jones J. Trends and Characteristics Associated With Human Papillomavirus Vaccination Uptake Among Men Who Have Sex With Men in the United States, 2014–2017. Sexually Transmitted Diseases 2019;46(7):465-473. [doi: 10.1097/olq.0000000000001008]
25. Checchi M, Mesher D, McCall M, Coukan F, Chau C, Mohammed H, et al. HPV vaccination of gay, bisexual and other men who have sex with men in sexual health and HIV clinics in England: vaccination uptake and attendances during the pilot phase. Sex Transm Infect 2019 Dec 26;95(8):608-613. [doi: 10.1136/sextrans-2018-053923] [Medline: 31028226]
26. Wang Z, Wang J, Fang Y, Gross DL, Wong MC, Wong EL, et al. Parental acceptability of HPV vaccination for boys and girls aged 9-13 years in China - A population-based study. Vaccine 2018 May 03;36(19):2657-2665. [doi: 10.1016/j.vaccine.2018.03.057] [Medline: 29606519]
27. Lau JTF, Wang Z, Kim JH, Lau M, Lai CHY, Mo PKH. Acceptability of HPV vaccines and associations with perceptions related to HPV and HPV vaccines among men who have sex with men in Hong Kong. PLoS One 2013 Feb 22;8(2):e57204 [FREE Full text] [doi: 10.1371/journal.pone.0057204] [Medline: 23451188]
28. Wang Z, Mo PK, Lau JT, Lau M, Lai CH. Acceptability of HPV vaccines and perceptions related to genital warts and penile/anal cancers among men who have sex with men in Hong Kong. Vaccine 2013 Sep 23;31(41):4675-4681. [doi: 10.1016/j.vaccine.2013.06.090] [Medline: 23845815]
29. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee A. Increasing Human Papillomavirus Vaccination Among Young Gay and Bisexual Men: A Randomized Pilot Trial of the Intervention. LGBT Health 2018 Jul;5(5):325-329. [doi: 10.1089/lgbt.2018.0059]
30. Reiter PL, Gower AL, Kiss DE, Malone MA, Katz ML, Bauermeister JA, et al. A Web-Based Human Papillomavirus Vaccination Intervention for Young Gay, Bisexual, and Other Men Who Have Sex With Men: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2020 Feb 24;9(2):e16294 [FREE Full text] [doi: 10.2196/16294] [Medline: 32130192]
31. Fontenot HB, White BP, Rosenberger JG, Lacasse H, Rutirasiri C, Mayer KH, et al. Mobile App Strategy to Facilitate Human Papillomavirus Vaccination Among Young Men Who Have Sex With Men: Pilot Intervention Study. J Med Internet Res 2020 Nov 04;22(11):e22878 [FREE Full text] [doi: 10.2196/22878] [Medline: 33146621]
32. Gerend M, Madkins K, Crosby S, Korpak AK, Phillips GL, Bass M, et al. Evaluation of a Text Messaging-Based Human Papillomavirus Vaccination Intervention for Young Sexual Minority Men: Results from a Pilot Randomized Controlled Trial. Ann Behav Med 2020 Sep 11:2020. [doi: 10.1093/abm/kaaa056] [Medline: 32914838]
33. Michie S, Johnston M, Francis J, Hardeman W, Eccle M. From theory to intervention: mapping theoretically derived behaviorual determinants to behaviour change techniques. Applied Psychology 2008;57:660-680. [doi: 10.1111/j.1464-0597.2008.00341.x]
34. Miller W, McLean SG, Duffin R, Lawal AO, Araujo JA, Shaw CA, et al. Diesel exhaust particulate increases the size and complexity of lesions in atherosclerotic mice. Part Fibre Toxicol 2013 Dec 11;10:61 [FREE Full text] [doi: 10.1186/1743-8977-10-61] [Medline: 24330719]
35. Complete Listing of Risk Reduction Evidence-based Behavioral Interventions. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/hiv/research/interventionresearch/compendium/rr/complete.html [accessed 2021-01-08]
36. Berg RC, Ross MW, Tikkanen R. The Effectiveness of MI4MSM: How Useful is Motivational Interviewing as an HIV Risk Prevention Program for Men who have Sex with Men? A Systematic Review. AIDS Education and Prevention 2011 Dec;23(6):533-549. [doi: 10.1521/aeap.2011.23.6.533]
37. Wang Z, Lau JTF, Ip M, Ho SPY, Mo PKH, Latkin C, et al. A Randomized Controlled Trial Evaluating Efficacy of Promoting a Home-Based HIV Self-Testing with Online Counseling on Increasing HIV Testing Among Men Who Have Sex with Men. AIDS Behav 2018 Jan 22;22(1):190-201. [doi: 10.1007/s10461-017-1887-2] [Medline: 28831616]
38. Factsheet: PRiSM - HIV prevalence and risk behavioral survey of men who have sex with men in Hong Kong. Centre for Health Protection. London: Centre for Health Protection, Department of Health; 2012 Aug. URL: https://www.aids.gov.hk/english/surveillance/sur_report/prism2011e.pdf [accessed 2021-01-08]
39. Janz N, Becker M. The Health Belief Model: a decade later. Health Education Quarterly. URL: http://heb.sagepub.com/content/11/41/41.short [accessed 2021-01-08]
40. Edmundson C, Heinsbroek E, Glass R, Hope V, Mohammed H, White M, et al. Sexualised drug use in the United Kingdom (UK): A review of the literature. Int J Drug Policy 2018 May;55:131-148. [doi: 10.1016/j.drugpo.2018.02.002] [Medline: 29625796]

41. Wang Z, Yang X, Mo PKH, Fang Y, Ip TKM, Lau JTF. Influence of Social Media on Sexualized Drug Use and Chemsex Among Chinese Men Who Have Sex With Men: Observational Prospective Cohort Study. J Med Internet Res 2020 Jul 24;22(7):e17894 [FREE Full text] [doi: 10.2196/17894] [Medline: 32706705]
42. Del Re AC, Maisel NC, Blodgett JC, Finney JW. Intention-to-treat analyses and missing data approaches in pharmacotherapy trials for alcohol use disorders. BMJ Open 2013 Nov 12;3(11):e003464 [FREE Full text] [doi: 10.1136/bmjopen-2013-003464] [Medline: 24227870]
43. Baron RM, Kenny DA. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 1986;51(6):1173-1182. [doi: 10.1037/0022-3514.51.6.1173]
44. The HPV Vaccine: Access and Use in the U.S. Henry J Kaiser Family Foundation. URL: http://files.kff.org/attachment/fact-sheet-the-hpv-vaccine-access-and-use-in-the-u-s [accessed 2021-01-08]
45. An evidence-based knowledge transfer program promoting HPV vaccination among men who have sex with men in Hong Kong (KPF18HLF22). Knowledge Transfer Initiatives. 2018. URL: https://tinyurl.com/1bk1j48p [accessed 2021-01-08]
46. Catania JA, Kegeles SM, Coates TJ. Towards an understanding of risk behavior: an AIDS risk reduction model (ARRM). Health Educ Q 1990 Sep 04;17(1):53-72. [doi: 10.1177/109019819001700107] [Medline: 2318652]
47. Prata N, Morris L, Mazive E, Vahidnia F, Stehr M. Relationship between HIV risk perception and condom use: Evidence from a population-based survey in Mozambique. Int Fam Plan Perspect 2006 Dec;32(4):192-200 [FREE Full text] [doi: 10.1363/3219206] [Medline: 17237016]
48. 96% of Hongkongers use their smartphone to go online everyday. South China Morning Post. 2013 Aug 21. URL: http://www.scmp.com/news/hong-kong/article/1298330/hong-kong-smartphone-use-doubles-two-years [accessed 2021-01-08]

Abbreviations

AOR: Adjusted odds ratios
CAI: condomless anal intercourse
CDC: Centers for Disease Control and Prevention
HBM: Health Belief Model
HPV: human papillomavirus
ITT: intention-to-treat
mHealth: mobile health
MI: motivational interviewing
MSM: men who have sex with men
NGO: nongovernmental organization
NRP: nonregular partners
OT: online tutorial
RCT: randomized controlled trial
RP: Regular male sex partners
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 16.06.20; peer-reviewed by S Song, M Threats; comments to author 12.11.20; revised version received 30.11.20; accepted 24.12.20; published 02.02.21.

Please cite as:
Wang Z, Lau JTF, Ip TKM, Yu Y, Fong F, Fang Y, Mo PKH
Two Web-Based and Theory-Based Interventions With and Without Brief Motivational Interviewing in the Promotion of Human Papillomavirus Vaccination Among Chinese Men Who Have Sex With Men: Randomized Controlled Trial
J Med Internet Res 2021;23(2):e21465
URL: https://www.jmir.org/2021/2/e21465
doi:10.2196/21465
PMID:33528372

Original Paper

Adoption of Digital Health Technologies in the Practice of Behavioral Health: Qualitative Case Study of Glucose Monitoring Technology

Suepattra G May[1*], MPH, PhD; Caroline Huber[2*], MPH; Meaghan Roach[1*], MPH; Jason Shafrin[1*], PhD; Wade Aubry[3*], MD; Darius Lakdawalla[4*], PhD; John M Kane[5,6*], MD; Felicia Forma[7*], BA

[1]PRECISIONheor, Los Angeles, CA, United States

[2]PRECISIONheor, New York, NY, United States

[3]Philip R Lee Institute for Health Policy Studies, University of California San Francisco, San Francisco, CA, United States

[4]University of Southern California, Los Angeles, CA, United States

[5]School of Medicine, Hofstra University, Hempstead, NY, United States

[6]Northwell Health, New York, NY, United States

[7]Otsuka Pharmaceutical Development & Commercialization Inc, Princeton, NJ, United States

[*]all authors contributed equally

Corresponding Author:
Suepattra G May, MPH, PhD
PRECISIONheor
11100 Santa Monica Blvd
Suite 500
Los Angeles, CA, 90025
United States
Phone: 1 310 984 7739
Email: suepattra.may-slater@precisionvh.com

Abstract

Background: Evaluation of patients with serious mental illness (SMI) relies largely on patient or caregiver self-reported symptoms. New digital technologies are being developed to better quantify the longitudinal symptomology of patients with SMI and facilitate disease management. However, as these new technologies become more widely available, psychiatrists may be uncertain about how to integrate them into daily practice. To better understand how digital tools might be integrated into the treatment of patients with SMI, this study examines a case study of a successful technology adoption by physicians: endocrinologists' adoption of digital glucometers.

Objective: This study aims to understand the key facilitators of and barriers to clinician and patient adoption of digital glucose monitoring technologies to identify lessons that may be applicable across other chronic diseases, including SMIs.

Methods: We conducted focus groups with practicing endocrinologists from 2 large metropolitan areas using a semistructured discussion guide designed to elicit perspectives of and experiences with technology adoption. The thematic analysis identified barriers to and facilitators of integrating digital glucometers into clinical practice. Participants also provided recommendations for integrating digital health technologies into clinical practice more broadly.

Results: A total of 10 endocrinologists were enrolled: 60% (6/10) male; a mean of 18.4 years in practice (SD 5.6); and 80% (8/10) working in a group practice setting. Participants stated that digital glucometers represented a significant change in the treatment paradigm for diabetes care and facilitated more effective care delivery and patient engagement. Barriers to the adoption of digital glucometers included lack of coverage, provider reimbursement, and data management support, as well as patient heterogeneity. Participant recommendations to increase the use of digital health technologies included expanding reimbursement for clinician time, streamlining data management processes, and customizing the technologies to patient needs.

Conclusions: Digital glucose monitoring technologies have facilitated more effective, individualized care delivery and have improved patient engagement and health outcomes. However, key challenges faced by the endocrinologists included lack of reimbursement for clinician time and nonstandardized data management across devices. Key recommendations that may be

relevant for other diseases include improved data analytics to quickly and accurately synthesize data for patient care management, streamlined software, and standardized metrics.

(J Med Internet Res 2021;23(2):e18119) doi:10.2196/18119

KEYWORDS

digital technology; chronic disease; blood glucose self-monitoring; diabetes self-management; real-time systems; mental illness; mobile phone

Introduction

Background

In 2017, 46.6 million US adults—nearly 1 in 5—were living with mental illness, ranging in severity from mild to severe [1]. Nearly a quarter of these adults (11.2 million) were afflicted with serious mental illness (SMI), defined as a diagnosis of mental illness (eg, schizophrenia and schizoaffective disorders, bipolar disorder, or psychosis) that is persistent, disabling, and requiring specialized psychiatric treatment [1,2]. These patients have higher morbidity and mortality rates than individuals without SMI because of untreated and/or preventable chronic or infectious diseases [3,4]. In recent years, digital health technologies have been increasingly developed and recognized as care management tools for patients with SMI [5].

Digital health technologies and digital health interventions such as smartphone apps and wearable technologies can transmit data to health care providers. These technologies have enormous potential to improve care, offering patients, providers, and caregivers greater access to and information about illness management, treatment monitoring and medication adherence, and outcomes for a breadth of conditions, including SMI [5-7].

Yet, despite these potential benefits, the uptake of digital health technology in SMI has been low, and behavioral health practitioners have been hesitant to fully embrace digital health interventions for a variety of reasons. These include limited evidence and perceptions of efficacy, aversion to change, concern about added workflow, lack of appropriate reimbursement, and an increasingly overwhelming variety of features and sensors [8-10]. In the area of mental illness, there are added concerns, including stigma [11], privacy, and uncertainty regarding patient acceptance. In many respects, mental illness may derive the most benefit from new digital technologies as patient clinical evaluations are typically based largely on patient or caregiver self-report of symptoms [12]. In addition, the measurement of domains such as sleep, activity, social interactions, and medication ingestion in real time can be highly valuable to patients, clinicians, and caregivers [13].

Digital health technology has great potential to improve SMI patient outcomes and disease management. However, an understanding of the potential barriers and facilitators to adoption is warranted. How specific digital health technologies have been implemented and ultimately adopted (or not) in other therapeutic areas can serve as useful case studies from which lessons for future digital health implementation in SMI may be drawn. For example, a recent review of digital health technologies to manage hypertension found that the settings and context in which interventions are introduced as well as the individuals involved influenced adoption [14]. A 2017 study examining the implementation of mobile apps for patients to support postsurgical rehabilitation in orthopedics found that digital literacy and the impact of the intervention on outcomes and workflow need to be accounted for [15]. Another study examining digital health interventions to improve medication adherence in diabetes and hypertension found no conclusive evidence of improved adherence with technologies that incorporated features such as interactive voice response or telemonitoring [16].

The literature documents the introduction of digital health technology interventions across the treatment spectrum, especially how these may be used to manage chronic illness outside of the clinic [17]. However, most of these digital health interventions are still relatively new. To better understand the adoption of digital health technology interventions over a longer period whereby the technology has also evolved, we consider the case of how providers have collected data on patient glycemic marker hemoglobin A_{1c} (HbA_{1c}) levels over time in diabetes care delivery. HbA_{1c} reflects the average blood glucose or blood sugar level in an individual over the past 3 months. In managing diabetes, physicians use HbA_{1c} to monitor how well a patient may be managing their diabetes. Historically, endocrinologists were reliant upon patient self-monitoring and patient or caregiver reports of glycemic events to inform diabetes management and treatment decision making. Over the past 20 years, digital health technologies, such as continuous glucose monitors (CGMs) and mobile glucose monitoring, support HbA_{1c} management by transmitting glucose levels to patients and clinicians and provide real-time insulin dose recommendations [18,19]. Each succession of new digital glucose monitoring technology has improved the quality, frequency, and relevance of monitoring glucose levels over time [20]. Such data have the potential to allow physicians to better understand a patient's condition and assess treatment response in real time. In diabetes care delivery, endocrinologists play an important role in promoting adoption of and adherence to digital glucometer devices and are also key users of the technology, as the devices provide data to be used in clinical decision making.

Objectives

This study sought to investigate the trajectory of technology adoption in the monitoring and treatment of another chronic disease—diabetes—to identify issues and facilitate improved adoption that may be relevant to the adoption of digital health interventions within the field of behavioral health. By adopting a qualitative case study approach, this study aims to understand the key facilitators to and barriers of clinician and patient

adoption of digital glucose monitoring technologies for diabetes and identify lessons learned that could be relevant for the treatment and monitoring of other chronic conditions, including SMI.

Methods

Study Design

Using qualitative data collection methods, we conducted focus groups with endocrinologists in 2 major metropolitan regions of the United States. Through focus group–guided discussions with these clinicians, we elicited perspectives of, and reports of experiences with, the implementation of digital glucose monitoring technologies in endocrinology practices over the past 2 decades and how such technologies impacted day-to-day clinical practice.

Participant Recruitment

We engaged a medical market research firm to identify and recruit potential participants from their panel of specialty health care providers. Bias was minimized by using probability sampling to select potential clinicians to receive a study invitation in 2 major metropolitan regions of interest in the United States. A convenience sample of endocrinologists treating patients with diabetes was invited by email to participate in the study. Physicians were eligible to participate if they (1) were board certified in endocrinology and currently practicing medicine, (2) had spent at least 80% of full-time equivalent in clinical practice treating patients with diabetes, and (3) had reported a minimum of 10% of their patient panel used a digital blood glucose monitoring device. We excluded physicians who had not been in clinical practice for a minimum of 10 years to ensure that participants could provide a relevant historical perspective on the evolution of digital glucose monitoring and its implementation in clinical practice.

Endocrinologists who met the basic study eligibility requirements were contacted by telephone and screened with a study-specific questionnaire to confirm eligibility and capture demographic data. Participants were compensated for study completion.

The Advarra Institutional Review Board (IRB) reviewed the study protocol, informed consent documents, and focus group discussion guides. The Advarra IRB determined that the study met the requirements for exemption from IRB oversight and granted a waiver of exemption from review.

Data Collection

Two 90-min focus groups were conducted in December 2018 in 2 large metropolitan regions in the United States. One member of the study team with doctoral training in qualitative research methods (SM) conducted the focus groups using a semistructured focus group discussion guide. A literature review was performed to inform the development of the discussion guide. Sample discussion guide questions are detailed below (Textbox 1). Focus groups were audio-recorded and transcribed verbatim. In addition to the focus group discussion, 2 members of the study team (CH and MR) recorded observational field notes.

Textbox 1. Sample discussion guide questions.

Perceptions of and experiences with diabetes management before digital glucose monitoring technologies

- Before the introduction of continuous glucose monitors (CGMs), what was the most optimal way to manage patients with diabetes, that is, how did you establish baseline glucose profiles, monitor glucose levels, etc?
- How did you first introduce self-monitoring or home glucometers to your patients?
- How was the transition from patient self-monitoring to continuous glucometers or technology?

Glucose monitoring technology adoption

- What prompted discussions on the need to adopt continuous glucose monitoring devices in your respective practices?
- What benefits did you anticipate and why?
- What potential disadvantages did you anticipate and why?
- How do you introduce the idea of continuous glucose monitoring technologies to patients?

Barriers of and facilitators to adoption and implementation

- What considerations, processes, or key personnel facilitated the decision to adopt glucose monitoring devices in your respective practices?
- How much personnel time or other resources have you or your practice invested into education and training?
- How much time do you or your staff devote to reviewing and/or discussing data from glucose monitoring technologies with patients?
- How do data from glucose monitoring technologies affect your clinical decision making?

Recommendations for strategies and approaches to facilitate adoption and uptake

- Looking back at the implementation, what do you think have been the most effective strategies and approaches for glucose monitoring technology adoption and uptake in your clinical practice?
- What challenges do you see to future adoption of glucose monitoring technologies in practice?
- What recommendations or advice do you have to increase glucose monitoring technology adoption?

Data Analysis

The constant comparative method was used to analyze the transcripts, facilitated by using MAXQDA qualitative data management software program (MAXQDA12; VERBI, GmbH) [21]. Our analytic approach used a comprehensive review of the transcripts incorporating the constant comparative method. Data analysis was consistent with the principles of thematic analysis outlined by Glaser and Strauss [21]. This approach to analysis involves identifying similarities and differences in the ways participants discussed their perspectives of and experiences with digital glucose monitoring technology adoption and is widely used in qualitative research [22]. The analysis involved 4 stages: (1) initial coding was undertaken by members of the project team with graduate training in qualitative data analysis (CH, SM, and MR). Each team member individually reviewed the data to develop a general overall impression of the data content, outline preliminary descriptive categories, and develop a preliminary coding framework. (2) Codes generated during preliminary coding were further refined and defined into descriptive coding categories, resulting in a final coding framework with a set of codes generated inductively and derived from the original interview questions. (3) The codes were applied to each focus group transcript. To assess the degree of coding consistency during this phase, one member of the project team (SM) reviewed all coding to ensure that the coders understood and applied the codes in accordance with the definitions in the coding dictionary. (4) Using the constant comparative method, concepts, themes, and patterns emerged from the data. These were integrated and refined into meaningful groupings by the team and based on consensus. The analysis provided insight into participants' experiences, motivations, and actions regarding treatment decision making related to glucose monitoring technology, yielding the salient themes presented in the results.

Results

Participant Sample

A total of 10 endocrinologists participated in the study. Most (6/10, 60%) of the sample was male, and the mean age was 48.6 (SD 5.9) years, with an average of 18.4 (SD 5.6) years since receiving their medical license. Of note, most of the sample (8/10, 80%) reported that more than half of their diabetic patient population used digital glucose monitoring technologies. Participant demographic characteristics are detailed below (Table 1).

Table 1. Participant demographics (N=10).

Variable	Participants, n (%)
Age (years), mean (SD)	48.6 (5.9)
Gender, n (%)	
Male	6 (60)
Female	4 (40)
Race, n (%)	
African American	0 (0)
Asian	5 (50)
Mixed or Other	0 (0)
Native Hawaiian or Pacific Islander	1 (10)
White	2 (20)
Missing or refused	2 (20)
Hispanic ethnicity, n (%)	
Hispanic	1 (10)
Not Hispanic	9 (90)
Missing or refused	0 (0)
Board certification in endocrinology, n (%)	10 (100)
Years since graduated medical school, mean (SD)	21.2 (5.8)
Years since obtaining medical license, mean (SD)	18.4 (5.6)
Primary practice setting, n (%)	
Academic medical center	1 (10)
Endocrinology group practice	1 (10)
Group model health maintenance organization	3 (30)
Multispecialty group practice	3 (30)
Solo practice	2 (20)
Years prescribing blood glucose monitors, n (%)	
2-5	1 (10)
6-10	5 (50)
More than 10	4 (40)
Proportion of patients with diabetes using a digital glucometer (%), n (%)	
Less than 10	0 (0)
10-25	2 (20)
25-50	0 (0)
50-75	2 (20)
More than 75	6 (60)

Key Findings

Digital Monitoring Technologies Changed the Treatment Paradigm in Diabetes Care Delivery

Before the availability of digital glucose monitoring technologies, clinical management of diabetes was predominantly reactive management by endocrinologists and largely based on laboratory test results, clinical signs, and patient self-reporting. Glucose testing required a waiting period, whether for urine strips (60 seconds) or blood finger pricks (a few seconds), which served as a barrier to patients regularly checking their levels [23].

Participants in this study noted that patient demand, clinical recommendations, and research, such as those presented at national professional conferences, drove the transition from laboratory tests and urine strips to digital glucose monitoring technologies. In addition, adoption was heavily influenced by Food and Drug Administration approval of digital glucose

monitoring devices and direct-to-consumer advertising; the former was important for validating physician trust in the technology, and the latter was influential in generating patient demand for digital glucose monitoring technologies.

Despite the demand, participants initially were apprehensive about the reliability (ie, accuracy of data) of glucose monitoring technologies. As broad adoption occurred and the devices became easier to use and more reliable, skepticism was overcome as described by one endocrinologist:

> *I mean, you're always skeptical when something new comes out. When the first pumps came out...people then didn't want to wear it, after wearing it for a little bit. And then it was uncomfortable. And so, if the person can't tolerate wearing something, it's just not going to work... That really dissuades people from wanting to start it, for one thing. So that was a concern. But as people get more tech savvy, it's less and less difficult. I mean, it's taken me a while but in the beginning, it was really hard. So it's definitely been an evolution.*

Digital Monitoring Has Improved Patient Outcomes and Facilitated More Effective Care Delivery and Patient Engagement

Before the introduction of digital glucose monitoring, efforts to reduce the risk of hypoglycemia were challenging for both endocrinologists and patients. Endocrinologists only had access to glimpses of a patient's glucose level at specific points of time, as opposed to a more complete picture of HbA_{1c} levels that are provided by continuous glucose monitoring technologies, especially information on whether blood sugar is trending down or trending up [20].

Participants in this study described the ways in which digital glucose monitoring technologies have provided a range of benefits for their patients. A primary benefit was the potential for the digital glucose monitoring devices to raise provider and patient awareness of glycemic events, thereby reducing the risk of disease-related morbidity and mortality.

Historically, endocrinologists depended on patient self-reporting and laboratory testing to inform diabetes management and treatment decision making. Today, diabetes technologies, such as continuous glucose monitoring and mobile glucose monitoring, can support HbA_{1c} management by transmitting glucose levels to patients and clinicians and provide real-time insulin dose recommendations [10]. Such technologies quantify glucose levels over time and provide insight into patterns and trends that allow the clinician to better understand a patient's diabetes and gauge how the patient is responding to treatment:

> *...in the last decade or so, a revolution happens in endocrinology...before then, we are just doing nothing because we have just a few insulin [options] ...And we're maybe lucky to get A1C of eight and you are happy by that. And patient was frustrated, doctor was frustrated because we cannot do much more. Just that's it. We have very much more advanced insulin and several type of class of medication now. And at the same time, we have very advanced technology. It's completely a game-changer now. Easily we can get A1C of seven or less, easily. And that's why doctors are more happy, patients are more happy.*

As endocrinologists are able to see accurate data in real time, they reported not having to rely as much on subjective patient or caregiver reports to aid in clinical decision making. The availability of continuous data also allowed providers to identify patterns or trends around glycemic events in their patients:

> *But truly, it's just more information. Prior to that...when we used NPH [neutral protamine Hagedorn insulin], everybody was getting low and we just never checked them. And they didn't check themselves. I feel like it's changed. It's going to change longevity, I mean, in terms of their lifespan. And it's going to make a huge difference. So I feel like if that's the standard of care now, just people will just jump on board.*

Participants reported that glucose monitoring technologies can also ease caregiver burden by allowing caregivers to monitor patients' glucose levels, which is especially important for patients susceptible to hypoglycemic events:

> *Now the technology…you can assign a surrogate. They could be assigned and they could be informed about blood sugars before even the patient recognizes. If you have a continuous glucose monitoring and you have those uploaders and you assign a couple of people, maybe a teacher or a parent. He's at work and he can alert the teacher. I see [inaudible] going down. You can do it remotely. So that's fabulous for usually type 1 diabetic patients.*

With the widespread adoption of digital glucose technologies, providers reported being able to customize and calibrate treatment—and more aggressively if need be—thereby improving clinician decision making and patient care delivery.

Digital Monitoring Technologies Have Introduced New Challenges for Clinical Care Delivery

The practice environment and associated availability of resources emerged as a key determinant of digital glucose monitoring technology adoption and use in clinical practice. Participants discussed that while the availability of data streamlined certain areas of diabetic patient management, these devices often generated more work and responsibilities than anticipated, such as data management and obtaining coverage approval and reimbursement.

Participants emphasized that both they and their patients experienced a significant burden and bureaucracy associated with insurance and reimbursement coverage of the devices and associated visits. Payer coverage of digital glucose monitoring technologies is highly varied, and most payers require preapproval for the device and all sensors, which may only be applicable for a limited duration (days or weeks at a time), requiring endocrinologists to repeatedly complete the burdensome process of obtaining preapproval.

Participants noted that there are very few billing codes for these types of digital health monitoring devices; therefore, providers may not receive compensation for the time they devote to educating patients and addressing general device-related management issues. In the following narrative, a participant describes the time he must spend obtaining insurance approval for broken equipment while alluding to the fact that he is not being reimbursed for the time he spends:

> *And then just the time because it's a durable medical piece of equipment, the time. Things break, the clip, the tube gets twisted...Just dealing with all of that takes a lot more time than before we had it. So I mean, patients will always be – they're short one tube being sent but the insurance is not going to allow them to have it. So I mean, then you're on the phone trying to make an exception.*

A related challenge is the lack of dedicated staff at some practices to support patients with digital glucose monitoring devices. Participants noted the inherent limitations of the devices, including technology failures, device repairs, and patient education and troubleshooting. In the absence of staff who can provide support with these devices, providers are left to deal with these issues directly, taking unplanned time from their busy schedules:

> *It's so important if you're lucky enough to have CDE [Certified Diabetes Educator] or a group of people to help you. That their only focus is the technology. So if you've got someone like that on your side, it makes a huge difference in terms of who's going to get started and who's going to get used to their equipment quicker and who's going to feel comfortable.*

Although the data generated by digital glucose monitoring technologies have improved the health of their patients, participants noted that the data generated are often not easily actionable for 2 reasons. First, the sheer amount of data the devices generate cause *data overload* for the providers. Second, the data are presented in a form that offers limited context about the patient and their diabetes management. These 2 issues make it challenging for providers to analyze and synthesize data during a short patient visit:

> *It takes 20 minutes just to download and the next patient is there. Two other patient already showed up. And this is time-consuming. Just adjusting the pump, doing the download.*

Participants noted an overall lack of funding and resources to support all of the structural challenges described: coverage and reimbursement, data management, and patient support.

Case Study Lessons: Recommendations and Opportunities for Increasing Uptake of Digital Health Technologies

In the focus group discussions, participants were explicitly asked what recommendations they would make to increase uptake of digital health technologies more generally, based on their experiences with digital glucose monitoring technologies. Participants in this study felt that if digital health technologies were going to have a meaningful and effective role in diabetes management, tangible steps needed to be taken by providers and payers to bridge gaps in care and reimbursement policy, which in turn would increase the access to and uptake of the technologies. These recommendations are not limited to the field of endocrinology and are applicable to the incorporation of digital health technologies in other therapeutic areas as well. These included the following recommendations:

- Expand coverage and streamline approval processes: Participants recommended expanding coverage for patients by following review and approval processes in place for specialty prescription drugs. Specifically, the endocrinologists believed that insurance review and coverage approvals for high-cost specialty drugs were far more streamlined than the review processes that payers require for interventions that fall under a device classification. From their perspective and experience, device approvals appear to invite more scrutiny by insurers than specialty drug approvals.
- Provide reimbursement for clinician time: Participants noted the substantial amount of time they or their staff spent supporting patients on glucose monitoring technology use and device maintenance. In addition, although endocrinologists appreciated the data that digital health technologies generated, they were not able to synthesize the vast amount of data downloaded during each office visit, which is often the only time they could bill for reviewing the data. To address both challenges, they recommended reimbursement for and appropriate billing codes to document the time devoted to device support and data review.
- Streamline software and data management across device platforms: Participants reported that a lack of standardization across digital glucose monitoring devices and corresponding platforms has proven burdensome for clinic staff and clinic operations. Specifically, endocrinologists indicated that standardized metrics and a streamlined transfer of digital monitoring technology data to their medical records would make clinic visits more efficient and provide the clinician more time to review data with the patient during the office visit. Several providers in this study indicated that it would be most beneficial to have the data from digital glucose monitoring technologies integrated with their electronic medical record software. In addition, in the absence of appropriate reimbursement noted above, they recommended that manufacturers develop algorithms to assist in analyzing and displaying clinically relevant and actionable information for physicians during the clinic visit.
- Digital health technologies should be customized to patient needs: Participants explicitly noted that digital health technologies should not be a one-size-fits-all approach to care management. Specifically, it is critical to assess patient receptivity to technology and tailor the digital health technology recommendation accordingly. Participants explained that some technologies are more advanced than others, such as those that synchronize data to a mobile phone app or that, in the case of glucose monitors, provide alarms if the patient's glucose levels are outside a given

range. Participants emphasized that some patients prefer more complex technologies, whereas others are better suited to simpler devices. Patient reluctance to try new or novel digital technology can often be countered by offering patients free trials, allowing the patient an opportunity to experience the technology without immediately committing to use.

Discussion

Principal Findings

Digital health technologies provide an opportunity to promote better treatment decision making and measure and monitor patient outcomes in real time. The objective of this study was to review the adoption of digital glucose monitoring as a case example of digital health penetration to identify lessons learned that may be applied to the adoption of digital health technologies in the realm of behavioral health. The endocrinologists participating in our study described the spectrum of benefits and challenges they experienced and detailed specific recommendations for how digital health monitoring technologies may be applied to the clinical management of other chronic conditions, including behavioral health. From their reported perspectives, digital health monitoring has improved care delivery by providing data about patients that can help to better monitor patient adherence and response to therapy and personalize treatment in a manner that best optimizes patient outcomes.

At the same time, participants in this study expressed concerns over the sheer volume of data that digital health technologies have generated and the urgent need to find solutions that can help to relieve physician burden and not impede the care they deliver. Such solutions include removing barriers to insurance approval, expanding categories of reimbursement, and developing software tools or algorithms that can help providers and patients navigate and understand data. Participants recommended synthesizing data into brief, tailored reports that can be used to inform clinical management decisions during an office visit. Separately, payers do not yet provide sufficient reimbursement for the provider time required to prescribe and manage digital technologies, which limits the opportunities for full integration of these devices into clinical care. Being able to quantify the long-term clinical and economic benefits of digital health technologies may be one mechanism by which reimbursement can be attained. Despite the challenges encountered, the endocrinologists in this study described how the continuous improvements in technology and the clear improvements in clinical outcomes for patients that they had observed over the years led to widespread adoption of the devices in clinical practice.

Limitations

Although this study has yielded suggestive findings, it is not without sampling-related limitations that merit consideration before interpreting the study and its implications. The findings represent the views of a small convenience sample of individuals who may be more vocal than the general endocrinologist population. However, similar themes were revealed and described in both focus groups, suggesting that the patterning of data may be similar in groups with similar characteristics. It should be noted that the study participants were self-selected in that they volunteered for the study. In addition, this study was limited to 10 endocrinologists practicing in 2 large metropolitan regions and may not represent experiences and perceptions of other types of providers who serve patients with diabetes (eg, primary care physicians or nurse practitioners) or providers who practice in different types of socioeconomic settings or geographic regions. Larger-scale work is needed to establish the generalizability of the findings.

Despite these limitations, this study provides preliminary qualitative data from a group of highly experienced clinicians on the key factors associated with successful implementation of digital monitoring technologies and strategies to mitigate and overcome potential barriers to uptake.

Comparison With Previous Studies

Previous studies have assessed barriers to and facilitators of the adoption of glucose monitoring technologies from the perspective of patients, their caregivers, and providers. Studies with patients and caregivers have found that patients who used glucose monitoring technologies perceived better control of their diabetes and improved confidence in controlling glucose levels, and patients were more likely to use glucose monitoring technologies if recommended by their clinician or family members [24,25]. Patient-reported barriers to use include high cost and limited reimbursement, device discomfort, disruptive alarms, unfamiliarity or distrust with the technology, and added burden to calibrate the device and understand and respond to data [24-29]. Several studies have also explored provider-level barriers and facilitators associated with the adoption of glucose monitoring technologies. In these studies, providers reported having insufficient time to interpret data, inadequate reimbursement or bureaucratic reimbursement processes, perceived issues with technology accuracy and security, and a lack of ancillary resources as barriers to adoption [30-32]. Although these studies provide a basis for understanding provider challenges and other barriers, surveys were employed to collect data, and the studies were not designed to explore, in depth, any nuances related to the adoption of technologies.

Before this study, specific gaps in the literature included (1) how providers learned about new technologies and the decision to adopt into clinical practice, (2) how providers decided to recommend technologies to their patients, and (3) how providers approached patient resistance to new digital health technologies. Although our study findings generally align with those of previous published studies, the specific processes by which glucose monitoring technologies have actually been adopted into clinical practice have been more fully elucidated in this study, including the recommendations that providers have for increasing future uptake of digital health technologies. This study also aligns with findings by Fonseca et al [33] who concluded that the adoption of CGM could be increased through standardized and tailored reporting coupled with expanded reimbursement to cover both the cost of the device (to patients) and the time spent advising patients on using the device (for providers). A recent report from the American Medical Association underscores this study's findings on digital health

technology adoption by physicians. Specifically, doctors are reported to assess the feasibility of using new digital tools based on 3 factors: effort (how seamless the integration of the innovation be), outcome (value to patients), and finance (how much it will benefit practice). The report also emphasized that any new tools or solutions must also address the issues of coding and coverage [34].

Lessons Learned and Implications for Future Research

Future research should consider the benefits of and challenges to the adoption of digital health technologies in other therapeutic areas, such as behavioral health. However, for meaningful and effective uptake to occur, tangible steps must be taken to substantively apply lessons learned. On the basis of our findings, some lessons learned that may help to bridge the gaps in care and increase adoption and adaptation of digital health technologies in the practice of behavioral health include:

- Generate robust evidence on how digital health in SMI affects physician, patient, and caregiver decision making: for instance, a large share of mobile health apps is dedicated to treating SMI [35]. For many of these apps, evidence demonstrating treatment efficacy and use for physicians is limited [36]. Some digital health technologies in the SMI space, however, have shown the potential to improve physician decision making and decrease health care costs [37,38]. In fact, new technologies include both pharmaceutical treatment and digital sensors [39].
- Expand coverage and streamline approval processes: our participants described a movement from reactive management and reliance on infrequent self-reporting to more timely and accurate assessment through the use of digital technologies. They also noted the value of improved provider and patient awareness and communication and the opportunity to reduce caregiver anxiety and burden, all of which can ultimately improve patient outcomes. However, the amount of time practitioners must devote to navigating administrative challenges for patients' digital monitoring technologies can impede uptake. If new technologies are shown to be cost-effective, access to digital technologies will be worth the additional cost.
- Provide reimbursement for clinician time and resources for staff training and patient support: providers in this study who had dedicated resources and personnel to support digital monitoring technologies in their practices observed positive impacts and higher long-term adherence and use. This included clinician review and synthesis of device data, providing support for securing coverage approvals, patient education and training, and continued device monitoring, maintenance, and patient follow-up. However, not all providers had access to dedicated personnel and resources. Reimbursement for clinician time and associated resources could facilitate the expanded adoption of digital health technologies. The cost of clinician time and staff training should be incorporated into any economic model of new treatment value.
- Customize digital health technology tools to the patient: recognizing that digital technologies are not a one-size-fits-all solution, it is critical that the technology fits each individual patient's needs and that incentives for providers are aligned with the heterogeneity of patients' particular preferences and needs, rather than clinic- or physician-level metrics.

Our findings suggest that there are substantial challenges to effectively incorporating digital technologies within the context of clinical care; however, these challenges are not insurmountable if identified and addressed up front. All of these considerations play a critical role in potentially improving the management of patients with SMI, where infrequent and subjective assessment, recall bias, problems with medication adherence, and enormous caregiver burden are significant unmet needs. Additional focus groups with other physician specialties will also be useful as they face similar decisions regarding if, when, and how to adopt new digital health technologies into real-world treatment of patients with SMI.

Conclusions

Endocrinologists reported that digital glucose monitoring technologies facilitated more effective, individualized care delivery and improved patient engagement and health outcomes in diabetes care delivery. However, challenges faced by clinicians included a lack of reimbursement for clinician time and nonstandardized data management across devices. Key recommendations that may be relevant for other diseases include improved data analytics to quickly and accurately synthesize data for patient care management, streamlined software, and standardized metrics. Future research should examine the extent to which these learnings are relevant to the adoption of digital health technologies for other therapeutic areas. Further research is also warranted to systematically quantify the challenges described in this study in a larger sample of clinicians. The results suggest strategies and areas of focus for stakeholders to mitigate barriers and bridge gaps in digital monitoring technology uptake.

Acknowledgments

The authors would like to thank Ainslie Hatch for her research guidance and the endocrinologists who participated in the study for their time and thoughtful insights.

Authors' Contributions

All authors have made a substantial contribution to the information submitted for publication and drafting or revising the paper and have read and approved the final paper.

Conflicts of Interest

SGM, CH, MR, and JS are employees of PRECISIONheor, a research consultancy to the health and life science industries, and JS holds equity in the Precision Medicine Group, its parent company. WA is a Scientific Advisor to PRECISIONheor and an employee of the University of California, San Francisco. DL formerly served as a Scientific Advisor to PRECISIONheor, and he holds equity in its parent company, Precision Medicine Group. He has received consulting fees, research support, or honoraria from the following sources: Amgen, Biogen, Genentech, GRAIL, Edwards Life Sciences, Novartis, Otsuka, and Pfizer. JMK has been a consultant to or received honoraria from Alkermes, Intracellular Therapies, Janssen, LB Pharma, Lundbeck, Merck, Neurocrine, Otsuka, Roche, Sunovion, Teva, Dainippon Sumitomo, Health Rhythms, Newron, Saladex, and Takeda. He is a shareholder of LB Pharma and the Vanguard Research Group. FF is an employee of Otsuka Pharmaceuticals and Commercialization, which provided funding to PRECISIONheor to conduct this study.

References

1. Mental Illness. The National Institute of Mental Health. Bethesda; 2019 Aug. URL: https://www.nimh.nih.gov/health/statistics/mental-illness.shtml [accessed 2020-12-21]
2. Bhugra D. The global prevalence of schizophrenia. PLoS Med 2005 May 31;2(5):e151 [FREE Full text] [doi: 10.1371/journal.pmed.0020151] [Medline: 15916460]
3. Novak PJ, Chen J, Ali M. Barriers to health care access among us adults with chronic conditions and co-ocurring serious psychological distress between 2011-2015. Medical Research Archives 2019 Jul 15;7(7) [FREE Full text] [doi: 10.18103/mra.v7i7.1947]
4. Hughes E, Bassi S, Gilbody S, Bland M, Martin F. Prevalence of HIV, hepatitis B, and hepatitis C in people with severe mental illness: a systematic review and meta-analysis. The Lancet Psychiatry 2016 Jan;3(1):40-48 [FREE Full text] [doi: 10.1016/s2215-0366(15)00357-0]
5. Batra S, Baker RA, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. MDER 2017 Oct 4;Volume 10:237-251 [FREE Full text] [doi: 10.2147/mder.s144158]
6. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
7. Naslund JA, Aschbrenner KA. Digital technology for health promotion: opportunities to address excess mortality in persons living with severe mental disorders. Evid Based Ment Health 2019 Feb 17;22(1):17-22 [FREE Full text] [doi: 10.1136/ebmental-2018-300034] [Medline: 30559332]
8. Felgner S, Ex P, Henschke C. Physicians' decision making on adoption of new technologies and role of coverage with evidence development: a qualitative study. Value Health 2018 Sep;21(9):1069-1076 [FREE Full text] [doi: 10.1016/j.jval.2018.03.006] [Medline: 30224111]
9. Nemec PB, Chan S. Behavioral health workforce development challenges in the digital health era. Psychiatr Rehabil J 2017 Sep;40(3):339-341 [FREE Full text] [doi: 10.1037/prj0000283] [Medline: 28891661]
10. Hendrikoff L, Kambeitz-Ilankovic L, Pryss R, Senner F, Falkai P, Pogarell O, et al. Prospective acceptance of distinct mobile mental health features in psychiatric patients and mental health professionals. J Psychiatr Res 2019 Feb;109:126-132 [FREE Full text] [doi: 10.1016/j.jpsychires.2018.11.025] [Medline: 30530207]
11. Rüsch N, Angermeyer MC, Corrigan PW. Mental illness stigma: concepts, consequences, and initiatives to reduce stigma. Eur Psychiatry 2005 Dec;20(8):529-539 [FREE Full text] [doi: 10.1016/j.eurpsy.2005.04.004] [Medline: 16171984]
12. Stahl SM, Morrissette DA, Faedda G, Fava M, Goldberg JF, Keck PE, et al. Guidelines for the recognition and management of mixed depression. CNS Spectr 2017 Feb 28;22(2):203-219 [FREE Full text] [doi: 10.1017/s1092852917000165]
13. Barnes AL, Murphy ME, Fowler CA, Rempfer MV. Health-related quality of life and overall life satisfaction in people with serious mental illness. Schizophr Res Treatment 2012;2012:245103 [FREE Full text] [doi: 10.1155/2012/245103] [Medline: 23213525]
14. Palacholla RS, Fischer N, Coleman A, Agboola S, Kirley K, Felsted J, et al. Provider- and patient-related barriers to and facilitators of digital health technology adoption for hypertension management: scoping review. JMIR Cardio 2019 Mar 26;3(1):e11951 [FREE Full text] [doi: 10.2196/11951] [Medline: 31758771]
15. Lau AY, Piper K, Bokor D, Martin P, Lau VS, Coiera E. Challenges during implementation of a patient-facing mobile app for surgical rehabilitation: feasibility study. JMIR Hum Factors 2017 Dec 07;4(4):e31 [FREE Full text] [doi: 10.2196/humanfactors.8096] [Medline: 29217504]
16. Conway CM, Kelechi TJ. Digital health for medication adherence in adult diabetes or hypertension: an integrative review. JMIR Diabetes 2017 Aug 16;2(2):e20 [FREE Full text] [doi: 10.2196/diabetes.8030] [Medline: 30291093]
17. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]

18. Klonoff DC. A review of continuous glucose monitoring technology. Diabetes Technol Ther 2005 Oct;7(5):770-775 [FREE Full text] [doi: 10.1089/dia.2005.7.770] [Medline: 16241880]
19. Dungan K, Reddy N, Verma N. Monitoring technologies- continuous glucose monitoring, mobile technology, biomarkers of glycemic control. In: Feingold KR, Anawalt B, Boyce A, editors. Endotext. South Dartmouth: MDText.com, Inc; Aug 16, 2020.
20. Vashist SK. Continuous glucose monitoring systems: a review. Diagnostics (Basel) 2013 Oct 29;3(4):385-412 [FREE Full text] [doi: 10.3390/diagnostics3040385] [Medline: 26824930]
21. Glaser BG, Strauss AL. The discovery of grounded theory. Sociology & Social Policy 1999;1:1-18 [FREE Full text] [doi: 10.4324/9780203793206-1]
22. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J of Quali Methods 2016 Nov 29;5(1):80-92 [FREE Full text] [doi: 10.1177/160940690600500107]
23. Clarke SF, Foster JR. A history of blood glucose meters and their role in self-monitoring of diabetes mellitus. British J of Biomed Sci 2018 Mar 22;69(2):83-93 [FREE Full text] [doi: 10.1080/09674845.2012.12002443]
24. Chua S, Ong WM, Ng CJ. Barriers and facilitators to self-monitoring of blood glucose in people with Type 2 diabetes using insulin: a qualitative study. PPA 2014 Feb 15;2014:237-246 [FREE Full text] [doi: 10.2147/ppa.s57567]
25. Pickup JC, Holloway MF, Samsi K. Real-time continuous glucose monitoring in Type 1 diabetes: a qualitative framework analysis of patient narratives. Diabetes Care 2015 Apr;38(4):544-550 [FREE Full text] [doi: 10.2337/dc14-1855] [Medline: 25552422]
26. Lawton J, Blackburn M, Allen J, Campbell F, Elleri D, Leelarathna L, et al. Patients' and caregivers' experiences of using continuous glucose monitoring to support diabetes self-management: qualitative study. BMC Endocr Disord 2018 Feb 20;18(1):12 [FREE Full text] [doi: 10.1186/s12902-018-0239-1] [Medline: 29458348]
27. Tanenbaum ML, Hanes SJ, Miller KM, Naranjo D, Bensen R, Hood KK. Diabetes device use in adults with Type 1 diabetes: barriers to uptake and potential intervention targets. Diabetes Care 2017 Feb;40(2):181-187 [FREE Full text] [doi: 10.2337/dc16-1536] [Medline: 27899489]
28. Naranjo D, Tanenbaum ML, Iturralde E, Hood KK. Diabetes technology: uptake, outcomes, barriers, and the intersection with distress. J Diabetes Sci Technol 2016 Jun 28;10(4):852-858 [FREE Full text] [doi: 10.1177/1932296816650900] [Medline: 27234809]
29. Tansey M, Laffel L, Cheng J, Beck R, Coffey J, Huang E, Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. Satisfaction with continuous glucose monitoring in adults and youths with Type 1 diabetes. Diabet Med 2011 Sep;28(9):1118-1122 [FREE Full text] [doi: 10.1111/j.1464-5491.2011.03368.x] [Medline: 21692844]
30. Derrick KM, Heptulla RA. Pediatric endocrinologists' experiences with continuous glucose monitors in children with Type 1 diabetes. J Diabetes Sci Technol 2018 Sep;12(5):1090-1091 [FREE Full text] [doi: 10.1177/1932296818776030] [Medline: 29808719]
31. Okazaki M, Kobayashi H, Ishii Y, Kanbori M, Yajima T. Real-world treatment patterns for golimumab and concomitant medications in Japanese rheumatoid arthritis patients. Rheumatol Ther 2018 Feb 22;5(1):185-201 [FREE Full text] [doi: 10.1007/s40744-018-0095-5] [Medline: 29470832]
32. Tanenbaum ML, Adams RN, Hanes SJ, Barley RC, Miller KM, Mulvaney SA, et al. Optimal use of diabetes devices: clinician perspectives on barriers and adherence to device use. J Diabetes Sci Technol 2017 Jan 10;11(3):484-492 [FREE Full text] [doi: 10.1177/1932296816688010] [Medline: 28745093]
33. Fonseca V, Grunberger G, Anhalt H, Bailey TS, Blevins T, Garg SK, Consensus Conference Writing Committee. Continuous glucose monitoring: a consensus conference of the American Association of Clinical Endocrinologists and American College of Endocrinology. Endocr Pract 2016 Aug;22(8):1008-1021. [doi: 10.4158/EP161392.CS] [Medline: 27214060]
34. Robezniaks A. Tomorrow's health tech must avoid the EHR's mistakes. Here's how. American Medical Association. 2019 Nov 21. URL: https://www.ama-assn.org/practice-management/digital/tomorrow-s-health-tech-must-avoid-ehr-s-mistakes-here-s-how [accessed 2020-12-21]
35. Anthes E. Mental health: there's an app for that. Nature 2016 Apr 07;532(7597):20-23 [FREE Full text] [doi: 10.1038/532020a] [Medline: 27078548]
36. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell appsvaluation of mental health app store quality claims. NPJ Digit Med 2019 Mar 22;18:- [FREE Full text]
37. Shafrin J, Forma F, Scherer E, Hatch A, Vytlacil E, Lakdawalla D. The cost of adherence mismeasurement in serious mental illness: a claims-based analysis. Am J Manag Care 2017 May 01;23(5):156-163 [FREE Full text] [Medline: 28810130]
38. Shafrin J, May SG, Shrestha A, Ruetsch C, Gerlanc N, Forma F, et al. Access to credible information on schizophrenia patients' medication adherence by prescribers can change their treatment strategies: evidence from an online survey of providers. Patient Prefer Adherence 2017;11:1071-1081 [FREE Full text] [doi: 10.2147/PPA.S135957] [Medline: 28721020]
39. Peters-Strickland T, Pestreich L, Hatch A, Rohatagi S, Baker RA, Docherty JP, et al. Usability of a novel digital medicine system in adults with schizophrenia treated with sensor-embedded tablets of aripiprazole. Neuropsychiatr Dis Treat 2016 Oct 11;12:2587-2594 [FREE Full text] [doi: 10.2147/NDT.S116029] [Medline: 27785036]

Abbreviations

CGM: continuous glucose monitor
HbA_{1c}: hemoglobin A_{1c}
IRB: Institutional Review Board
SMI: serious mental illness

Edited by G Eysenbach; submitted 04.02.20; peer-reviewed by B Aylward, S Kim, M Ameen; comments to author 29.06.20; revised version received 05.08.20; accepted 23.11.20; published 03.02.21.

Please cite as:
May SG, Huber C, Roach M, Shafrin J, Aubry W, Lakdawalla D, Kane JM, Forma F
Adoption of Digital Health Technologies in the Practice of Behavioral Health: Qualitative Case Study of Glucose Monitoring Technology
J Med Internet Res 2021;23(2):e18119
URL: https://www.jmir.org/2021/2/e18119
doi:10.2196/18119
PMID:33533725

Review

Identification of Behavior Change Techniques From Successful Web-Based Interventions Targeting Alcohol Consumption, Binge Eating, and Gambling: Systematic Review

Gabrielle Humphreys[1], BSc, MSc; Rebecca Evans[1], BSc, MSc; Harriet Makin[1], BSc, MSc; Richard Cooke[1], BSc, MSc, PhD; Andrew Jones[1], BSc, MSc, PhD

School of Psychology, University of Liverpool, Liverpool, United Kingdom

Corresponding Author:
Gabrielle Humphreys, BSc, MSc
School of Psychology
University of Liverpool
Eleanor Rathbone Building, L69 7ZT
Liverpool
United Kingdom
Phone: 44 07311550101
Email: gabrielle.humphreys@liverpool.ac.uk

Abstract

Background: Web-based interventions are thought to overcome barriers to treatment, such as accessibility and geographical location, which can undermine the effectiveness of traditional face-to-face interventions. Owing to these features, researchers are increasingly testing the efficacy of web-based interventions as ways to reduce alcohol misuse, binge eating, and gambling. However, many web-based interventions have poorly defined mechanisms of action; therefore, it is often uncertain how they propose to bring about behavior change.

Objective: This systematic review aims to identify effective behavior change techniques (BCTs) present in web-based interventions aimed at reducing alcohol consumption, binge eating, or gambling.

Methods: This systematic review covered research conducted in the last 20 years. Inclusion criteria for interventions were web-based administration; targeting alcohol use, binge eating, and/or gambling; and reporting on baseline and postintervention measures of behavior. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines were followed. We coded intervention effectiveness, study quality, and BCTs present in the interventions.

Results: Following removal of 4152 ineligible articles, 45 were included in the review: 32 (71%) targeted alcohol misuse, 6 (13%) targeted binge eating, and 7 (16%) targeted gambling. In total, 5 frequency counts were performed to identify the most commonly used BCTs: all studies, effective interventions, high-quality studies at 2 thresholds, and both high quality and effective studies. The results obtained from this were integrated to identify 7 BCTs. These 7 BCTs were problem solving, feedback on behavior, self-monitoring of behavior, self-monitoring of outcomes, instruction on how to perform a behavior, information about social and health consequences, and social comparison. A total of 4 BCTs were found in all frequency counts: feedback on behavior, self-monitoring of behavior, instruction on how to perform a behavior, and social comparison. Self-monitoring of outcomes of behavior was found in 3 of the 5 frequency counts, problem solving was found in 2 frequency counts, and information about social and health consequences was found in 1 frequency count.

Conclusions: This systematic review identified 7 of the most frequently used BCTs used in web-based interventions focused on alcohol misuse, binge eating, and gambling. These results can inform the development of evidence-based eHealth interventions that have the potential to lead to effective, positive behavior changes in all 3 areas.

(J Med Internet Res 2021;23(2):e22694) doi:10.2196/22694

KEYWORDS

systematic review; web-based intervention; behavior and behavior mechanism; behavior change technique; alcohol consumption; binge eating; gambling

Introduction

Background

Alcohol misuse, binge eating, and gambling are increasing in prevalence in Western countries [1-3]. Importantly, these behaviors often occur independently of a formal diagnosis of a relevant behavioral or substance use disorder, or an eating disorder in subclinical populations. This may be due to the general acceptability of these behaviors, compared with other addictive behaviors [4]. These attitudes may be explained by the legality of these behaviors and substances and their widespread accessibility. In turn, this may normalize these behaviors, something that may be problematic in individuals displaying concerning behavior in these areas. As such, many otherwise *healthy* individuals will indulge in one or perhaps all of these behaviors over the course of their lifetime. Furthermore, these behaviors can cause considerable mental and physical harm to an individual (eg, liver disease, pancreatic disease, impaired cognition [5]) and may exacerbate mental or physical health disorders. Highlighted in the UK National Health Services' Long Term Plan (2019) [6], the effect of these behaviors will also disperse into wider society, placing further strain on health care systems [7].

Despite the introduction of various regulations, alcohol remains a highly accessible and an attractive pastime for most in the western world [8]. Recent data reported that 29% of adults' alcohol consumption in the United Kingdom is hazardous [9], categorized as typically drinking over 14 units of alcohol a week. This remains to be a global problem, with 39.5% of drinkers reporting heavy episodic drinking globally [10].

Similar to alcohol consumption, disordered eating (or binge eating) has been normalized by society [11,12]. Binge eating is the consumption of a larger amount of food than what is expected of an individual, in a fixed period, typically accompanied by a sense of lack of control [13]. This behavior is typically used as a coping mechanism in response to adverse negative events or stimuli [14], which may be conscious or unconscious (thus, the term *emotional eating* is often used). Binge eating is linked to classified eating disorders, including binge eating disorder, bulimia nervosa, anorexia nervosa, and eating disorders not otherwise specified [14]. However, this act may also be present in subclinical populations without a diagnosis. Over the last decade, hospital admissions for both eating disorders and mental health disorders have risen [15,16]. Technology may have contributed to the increase in disordered eating due to the promotion of unhealthy behaviors (eg, restrictive dieting, which leads to binge eating [17,18] and mukbang trends [19]) on web-based platforms and the 24/7 access to this content. Advances in technology may have also contributed to the increased prevalence of gambling and problem gambling [20]. Due to the extensive growth in web-based gambling, the opportunity to gamble is no longer limited to physical gambling establishments.

Theoretical models and empirical evidence suggest that alcohol misuse, binge eating, and gambling may share similar psychological mechanisms of action [21]. For example, individuals with increased impulsivity engage in greater and more frequent health and financial risk behaviors, such as gambling, binge eating, and binge drinking [22]. Similarly, alcohol misuse, binge eating, and gambling have all been characterized as compulsive behaviors, with the initiation of each thought to be driven by relevant behavior-related and emotive cues [23,24], suggesting cue-controlled behavior. The identification of common underlying behavioral mechanisms suggests that interventions targeting these broad behaviors, rather than tailoring them to specific symptoms, may be beneficial for a greater number of individuals [25,26].

Many individuals actively avoid seeking a formal diagnosis of their alcohol misuse, binge eating, or gambling due to negative self and social stigma [27-29]. This is despite many individuals recognizing the negative impact of these addictions on their quality of life (including their physical health, mental distress, and financial burden [30,31]). Furthermore, individuals may not seek traditional face-to-face treatments due to geographical restrictions, availability, or time constraints [32]. This has resulted in a considerable treatment gap [33], with many people in need not seeking or receiving treatment.

eHealth interventions are able to bridge this treatment gap by allowing anonymity, overcoming location barriers, and providing *any-time* accessibility [32,34]. There have been several successful interventions for each behavior. Meta-analyses of alcohol, gambling, and binge eating behavior have concluded that eHealth interventions are efficacious, with a moderate-to-large effect size [35] found for binge eating [36] and moderate effect sizes for alcohol [37] and gambling [38]. Despite these findings, many web-based intervention assessments lack vigorous testing methods, such as randomized controlled trials (RCTs) to determine their behavior change effectiveness [39]. Alternatively, if the effectiveness of the web-based intervention is tested, the reasons for this effectiveness, such as the behavior change techniques (BCTs) used, may not be explored [40]. This can be observed in many recent meta-analyses. Instead of this top-down approach, interventions should be developed following a bottom-up approach and the values of the Experimental Medicine Approach [41,42]. This approach focuses on examining the core mechanisms that are behind a behavior and holds particular use in intervention development, placing emphasis on the importance of identifying and targeting the malleable psychological processes or surroundings of an individual for behavior change [42].

The BCT taxonomy provides an overview of methods for behavior change and their hierarchical structure [43]. By looking at the underpinning mechanisms behind a behavior, an appropriate BCT, or more likely, a combination of BCTs can be identified to target a behavior effectively. Using this theory-based procedure to design an intervention means that optimum results are achieved for both the intervention users and business owners due to increased user satisfaction. eHealth interventions are often favored by businesses because of their cost-effectiveness in comparison with traditional face-to-face services. This cost-effectiveness increases dramatically when the transdiagnostic capability of a web-based intervention—when one treatment is used to target multiple behaviors—is considered [44]. Web-based interventions are

capable of being transdiagnostic, as their content is often open for interpretation, focusing on user reflection and self-monitoring of behavior.

Research has confirmed the efficacy of many eHealth interventions in changing addictive behaviors such as binge eating, gambling, and alcohol behavior [36-38]. Some potential moderators have been discussed. For example, gambling intervention effectiveness was found to be moderated by gender, with greater success in male participants [38]. In addition, group components of gambling interventions were found to have long-lasting behavioral effects [38]. In alcohol interventions, the delivery type was explored [37]. Greater success in reduced alcohol consumption was reported in interventions that provided supplementary alcohol-related materials, which were not *commercially available*, possibly implying the importance of in-depth, systematic intervention design for specific problems [37]. Motivation was also considered, with particular attention paid to the impact of eating intervention effectiveness on this, where over a quarter of studies used participants who were seeking weight loss through diet and exercise or bariatric surgery [36]. However, no specific consideration of BCTs were found in these studies. This is still common in several papers on behavior change [45,46]. This holds significant because of the concerns that many eHealth interventions are not developed on the basis of best practices in behavior change. Often, interventions are developed favoring intuition over research findings or those that claim to be evidence based are actually *evidence inspired* at best due to companies' preconceived ideas [46,47]. This is detrimental not only to intervention users and those delivering them but also for future research in behavior change and intervention development, as it leads to studies reporting effectiveness in behavior change but not identifying the specific mechanism behind this. The same applies to many evidence-based papers, with it being common for journal articles to lack detail on an intervention to allow for BCT coding [45], which will limit our current understanding of BCTs and mechanisms of behavior. Although it is a relatively new concept, the BCT taxonomy [43] provides a common language among behavior change papers, allowing effective synthesis of evidence-based papers to enhance behavior change knowledge.

The use of this taxonomy has led to valuable findings. Conducting a review on a single behavior type allows knowledge to be summarized in this specific area. This allows BCTs to be identified as effective for this topic, which holds value because BCT effectiveness is dependent on context, such as the type of behavior targeted [47]. Existing research has highlighted promising BCTs for alcohol, gambling, and binge eating. For example, Crane et al [48] identified 12 BCTs that were deemed as promising for behavior change, including self-monitoring of behavior, goal setting (behavior), feedback on behavior, and behavior substitution. Similar studies have been conducted on gambling, highlighting feedback on behavior, goal setting, social comparison, and exposure as useful BCT techniques to reduce gambling behavior [49]. A recent meta-analysis of BCTs on healthy eating interventions, which aimed to reduce unhealthy eating behaviors such as binge eating, similarly identified self-monitoring of both behavior and outcomes of behavior as a critical BCT for behavior change effectiveness [50]. A synthesis on BCTs can also be performed on multiple behaviors to inform the development of transdiagnostic interventions. However, there are no existing reviews examining which BCTs may be effective for binge eating, gambling, and alcohol consumption together.

Objectives

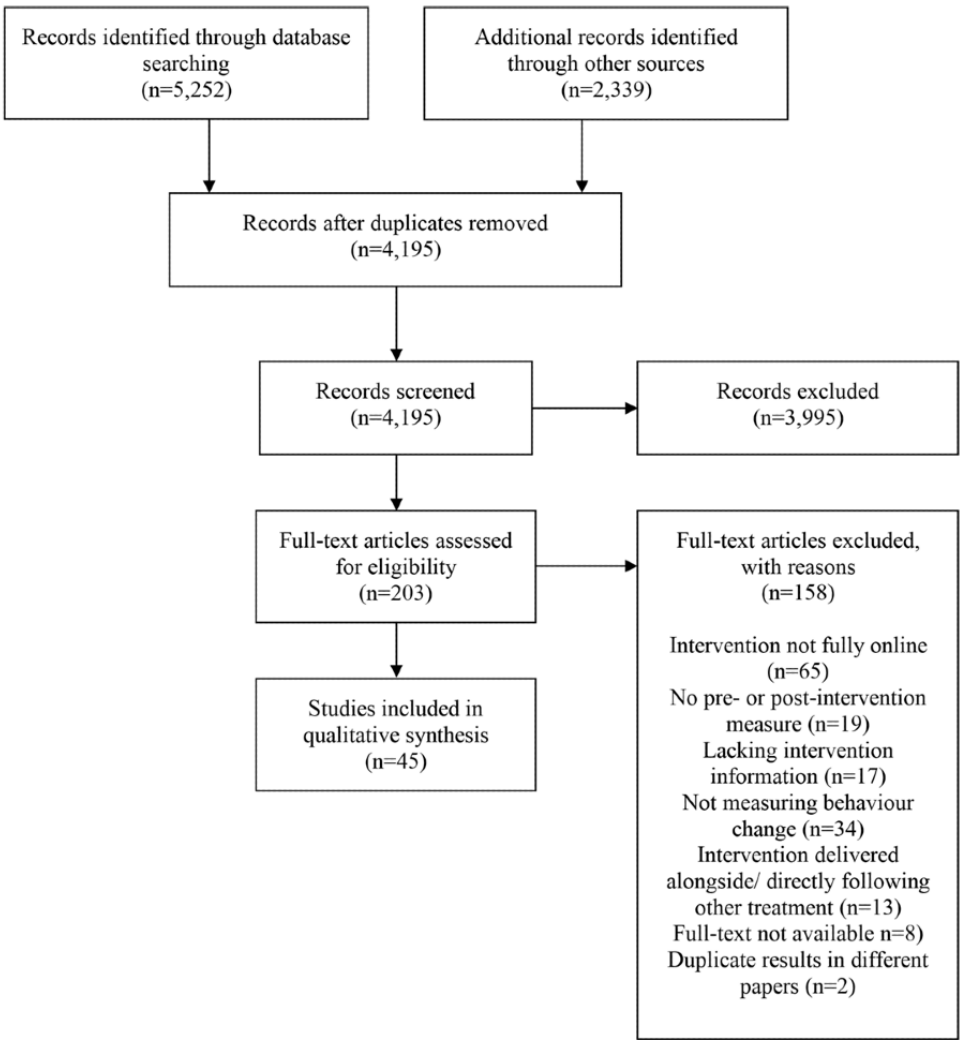
The aim of this systematic review is to (1) identify web-based interventions targeting a reduction in alcohol consumption, binge eating, and gambling behavior; we focused on these 3 behaviors due to their increasing prevalence in subclinical populations, similar underlying psychological mechanisms, and the widespread accessibility to materials for these behaviors; (2) determine the effectiveness of these interventions; and (3) identify the BCTs used in these interventions, examining the similarities in techniques to inform the development of a future transdiagnostic intervention.

Methods

Literature Search

Scoping searches were conducted in November 2019 using several electronic databases. Following these searches, we refined our search criteria and registered the protocol for this study on the Open Science Framework [51]. Full searches were conducted in December 2019 on PsycINFO, PubMed, and Scopus following the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines (Figure 1). Search terms can be found in Multimedia Appendix 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) flowchart followed in this study.



Eligibility Criteria

Relevant articles were identified by a single author (GH). A second coder (RE) independently assessed the eligibility of the selected papers. There were no major disagreements between the authors. To be eligible for inclusion in the review, studies had to be either case controlled or RCTs that examined the effectiveness of a web-based intervention. Intervention effectiveness was determined as measured behavior change (eg, reduction in units of alcohol consumed) rather than user satisfaction. Eligible studies had to examine interventions to reduce alcohol use, binge eating, or gambling, although these outcomes did not have to be the primary outcome measure of the study. Other inclusion criteria were as follows: measurements reported at baseline and immediately postintervention for the relevant behavior change, *either* sufficient detail of the intervention content in the paper *or* a direct link to the study protocol to allow for BCTs to be coded, and research published in the past 20 years with a human-only sample. Exclusion criteria included interventions that were not entirely delivered on the web, which means that any intervention with a face-to-face element was not eligible. Interventions not delivered independently of other behavior change methods (eg, an eating intervention in individuals who had previously undergone a gastric bypass) and interventions that aimed to maintain, rather than change, behavior were not included. Similarly, interventions aimed at preventing the onset of problematic behavior were not included. However, if such interventions aimed to prevent the worsening of behavior and measured this behavior, rather than reported user intentions, these studies were eligible [52]. Data collected must have been from a validated measure (eg, Alcohol Use Disorder Identification Test [AUDIT-C]) [53] or self-reported retrospective recall of quantity or frequency of behaviors (eg, Alcohol Timeline FollowBack [54]). Measurements had to encapsulate behavior change rather than acceptability, feasibility, or user satisfaction.

Data Extraction

Data were extracted from eligible studies by a single author (GH) using a data extraction form. Variables in the data extraction form included participant characteristics (including details of control groups), intervention characteristics, intervention effectiveness, and BCTs used in the intervention (Multimedia Appendices 2-4 [42,52,55-99]). BCTs were coded using the BCT taxonomy version 1 [43]. Some participants (eg, Crombie et al [55] and Pederson et al [56]) completed unrelated tasks as a comparator, referred to as an active control group.

Owing to their inability to eliminate placebo effects [100], if an active control group was used and adequate information was provided, the authors coded both interventions for BCTs. If a no-treatment control group was used, typically in the form of a waitlist condition or assessment-only condition, the intervention was not coded. The coding of BCTs was completed independently by 2 authors (GH and HM), with any disagreements resolved. There were no major disagreements with a Cohen κ score of 0.83. The authors were contacted if relevant data were missing from eligible papers. After the BCTs were coded, frequency counts were performed to examine the prevalence of BCTs. Data extracted for eligible studies [52,57-98] can be located in Multimedia Appendices 2-4. The BCT Periodic Table by Armitage et al [99] was used for labeling BCTs, rather than the BCT taxonomy v1 numbers [43], for ease of reading.

Quality Assessment

Risk of bias was assessed using the Office of Health Assessment and Translation (OHAT) Risk of Bias Rating Tool [101] selected due to its use in health-related studies and RCTs. The areas considered were as follows: condition randomization and concealment (where applicable), the appropriateness of control groups (if used), attrition rates, reporting bias, and the reliability and validity of measures. Item 5 from the OHAT tool was removed as it was only applicable to animal experiments, which were not included in this review. The number of items completed was dependent on study design, with a total score of either 30 (RCTs) or 18 (case controlled). Questions were answered using a 4-point scale ranging from *definitely low risk of bias* to *definitely high risk of bias*. Direct evidence was required for the scoring of definite high or low bias. Probable risk was scored if evidence was missing or indirect. A score of 3 coincided with a *definitely low risk of bias* and zero with a *definitely high risk of bias*. High quality was indicated by a score of over 70%, and medium quality was indicated by a score over 50% [101]. Despite the recommendation of 70%, the existing literature has also used 80% as a high-quality threshold. Results using both thresholds were compared.

Data Synthesis

An overall frequency count of BCTs used in the interventions was performed. BCTs were also counted when ineffective and when low-quality studies (<70% OHAT score and <80% OHAT score) were excluded, with results compared. The effectiveness of an intervention was determined by a statistically significant effect (P<.05) on behavior change in drinking, gambling, or eating. If a paper reported mixed findings regarding behavior change effectiveness, we used the most relevant variable as an indicator of effectiveness. The variable tended to be the overall effect reported or the most relevant measure of alcohol, binge eating, or gambling behavior if the intervention's primary aims were not directly to promote this change. For example, an intervention that aimed to promote weight loss by targeting binge eating was eligible (Lyzwinski et al [95]). Similarly, an intervention with the primary aim to reduce drunk driving was eligible as it also targeted alcohol consumption (Bingham et al [62]). Studies that reported a statistically significant reduction in the relevant behavior change measure were therefore included in the frequency count of *effective interventions*. Similarly, those scoring over 70% on the OHAT risk of bias tool were included in the frequency count of *high-quality papers*. Finally, those that reported both a statistically significant effect in the relevant behavior change measure and a score of over 70% on the OHAT risk of bias tool were included in the *effective and high-quality papers*. The 5 most commonly used BCTs were reported for each study (Table 1); if BCT frequencies were equal, over 5 were reported in the results.

Table 1. A summary table of the behavior change techniques identified in frequency counts.

Frequency count key	BCT[a] item						
	PS[b]	FOB[c]	SB[d]	SOB[e]	IPB[f]	ISEC[g]	SC[h]
Total[i]	✓[j]	✓	✓	—[k]	✓	✓	✓
Effective[l]	—	✓	✓	✓	✓	—	✓
>70% OHAT[m]	✓	✓	✓	—	✓	—	✓
>80% OHAT[n]	—	✓	✓	✓	✓	—	✓
>70% and effective[o]	—	✓	✓	✓	✓	—	✓

[a]BCT: behavior change technique.

[b]PS: problem solving.

[c]FOB: feedback on behavior.

[d]SB: self-monitoring of behavior.

[e]SOB: self-monitoring of outcomes of behavior.

[f]IPB: instruction on how to perform the behavior.

[g]ISEC: information about social and environmental consequences.

[h]SC: social comparison.

[i]Total: all eligible papers.

[j]✓: Presence of behavior change technique.

[k]—: Absence of behavior change technique.

[l]Effective: all papers that concluded a significant reduction in the specific target behavior measure from pre to postintervention.

[m]All papers deemed as high quality, which scored over 70% on the Office of Health Assessment and Translation (OHAT) risk of bias tool.

[n]All papers that scored over 80% on the OHAT risk of bias tool.

[o]All papers that scored over 70% on the OHAT risk of bias tool and were deemed as effective in target behavior reduction.

Results

Searches and Study Information

The search yielded a total of 5252 papers: 2672 alcohol papers, 1670 gambling papers, and 910 binge eating papers. After duplicates were removed, 4152 articles remained. Papers were then screened via title, abstract, and full text. Of those eligible papers, references were also examined (n=2339). This screening identified 2 papers that were eligible for the study. In total, 45 studies were identified as eligible for this review: 32 studies (71%) describing interventions for alcohol use, 7 (16%) describing interventions for gambling, and 6 (13%) describing interventions for binge eating.

Most eligible papers (37/45; 82%) were RCTs with a waitlist control group. The mean duration of intervention was 6 weeks. Interventions ranged from less than 1 hour to 1 year. A computer was the most frequently used intervention device. Participants were typically from at-risk groups or displayed concerning behavior (Multimedia Appendices 2-4). Pre- and postintervention behavior was most commonly assessed using the AUDIT-C for alcohol, the South Oaks Gambling Screen for gambling, and the Eating Disorder Examination Questionnaire for binge eating.

The most commonly used BCT across all 45 papers was self-monitoring of behavior (BCT item 2.3), which was present in 39/45 (87%) of interventions. Feedback on behavior (BCT item 2.2; 30/45, 67%), social comparison (BCT item 6.2; 27/45, 60%), information about social and environmental consequences (BCT item 5.3; 26/45, 58%), instruction on how to perform a behavior (BCT item 4.1; 25/45, 56%), and problem solving (BCT item 1.2; 25/45, 56%) were other commonly used techniques. Data on BCT frequency counts can be found in Multimedia Appendices 2-4.

Effective Interventions

In total, 66% (21/32) of alcohol interventions, 83% (5/6) of eating behavior interventions, and 43% (3/7) of gambling interventions were effective. After excluding unsuccessful interventions, 64% (29/45) of the eligible interventions remained. In total, 21 interventions targeted alcohol behavior (72%), 5 targeted binge eating (17%), and 3 targeted gambling (10%). A frequency count of effective-only papers revealed that the most commonly used BCTs remained self-monitoring of behavior (BCT item 2.3), feedback on behavior (BCT item 2.2), social comparison (BCT item 6.2), and instruction on how to perform a behavior (4.1). Information about social and environmental consequences (BCT item 5.3) was replaced by BCT item 2.4 (self-monitoring of outcomes of behavior). Thus, the BCTs used in effective studies were the same as those used most frequently across all studies.

High-Quality Papers

In total, 56% (25/45) of papers were rated as having high quality (OHAT>70%). Of these 25 papers, 18 (72%) papers targeted alcohol misuse, 4 (16%) targeted gambling, and 3 (11%) targeted binge eating. When examining only high-quality studies, the most frequent BCTs used were self-monitoring of behavior (BCT item 2.3), instruction on how to perform a behavior (BCT

item 4.1), feedback on behavior (BCT item 2.2), and social comparison (BCT item 6.2). Problem solving (BCT item 1.2) was also identified as a commonly used BCT in high-quality papers.

When the threshold for a high-quality study was increased (OHAT>80%), 9 papers retained this classification: 6 targeting alcohol abuse and 3 targeting gambling. No binge eating studies met this criterion. Self-monitoring of behavior (BCT item 2.3), feedback on behavior (BCT item 2.2), social comparison (BCT item 6.2), and instruction on how to perform a behavior (BCT item 4.1) remained to be the most commonly used BCTs. Problem solving (BCT item 1.2) was no longer within the top 5 most frequently used BCTs. Self-monitoring of outcomes of behavior replaced problem solving (BCT item 2.4), meaning results were consistent with the effective-only frequency count.

Effective and High-Quality Papers

In total, 16 of 45 papers (36%) were classified as being effective and of high quality (>70% OHAT score). Of these, 10 (63%) focused on alcohol misuse, 3 (19%) on binge eating, and 3 (19%) on gambling. The most commonly used BCTs in these papers were self-monitoring of behavior (BCT item 2.3), feedback on behavior (BCT item 2.2), self-monitoring of outcomes of behavior (BCT item 2.4), instruction on how to perform a behavior (BCT item 4.1), and social comparison (BCT item 6.2). This is consistent with the findings of the effective-only frequency count and the OHAT score over an 80% frequency count.

Common BCTs Among Frequency Counts

In total, 7 BCTs were identified as commonly used within the frequency counts. These were problem solving (BCT item 1.2), feedback on behavior (BCT item 2.2), self-monitoring of behavior (BCT item 2.3), self-monitoring of outcomes of behavior (BCT item 2.4), instruction on how to perform a behavior (BCT item 4.1), information about social and environmental consequences (BCT item 5.3), and social comparison (BCT item 6.2). Within the 5 frequency counts performed, the techniques feedback on behavior (BCT item 2.2), self-monitoring of behavior (BCT item 2.3), instruction on how to perform a behavior (BCT item 4.1), and social comparison (BCT item 6.2) were present in all counts. Self-monitoring of outcomes of behavior (BCT item 2.4) was identified in the majority of frequency counts (3). Problem solving (BCT item 1.2) was present in 2 frequency counts, and information about social and environmental consequences (BCT item 5.3) was present in 1 frequency count.

Discussion

Principal Findings

This systematic review identified 7 commonly used BCTs in web-based alcohol, gambling, and binge eating eHealth interventions: problem solving, feedback on behavior, self-monitoring of behavior, self-monitoring of outcomes of behavior, instruction on how to perform a behavior, information about social and environmental consequences, and social comparison. Although the most frequent BCTs used in papers varied when intervention effectiveness and study quality were taken into account, this variation was minor, with feedback on behavior, self-monitoring of outcomes, instruction on how to perform a behavior, and social comparison present in all frequency counts. This suggests that researchers testing interventions for these behaviors use similar approaches. Self-monitoring of outcomes was present in 3 out of 5 frequency counts, problem solving was present in 2 frequency counts, and information about social and environmental consequences was present in 1 frequency count. Attention should be paid to items feedback on behavior, self-monitoring of behavior, self-monitoring of outcomes, instruction on how to perform a behavior, and social comparison, which were the frequency count results for both high-quality and effective studies. Most papers were of high quality and were RCTs. However, when looking at the breakdown of studies in each behavior type, these were very imbalanced, with 71% targeting alcohol consumption, 16% gambling, and 13% binge eating. Study effectiveness varied within each behavior type, meaning that the studies included in effective frequency counts were not equally split (72% alcohol, 17% binge eating, and 10% gambling). However, this was expected due to the prevalence of interest in alcohol research, and thus, more published studies in this area. Despite highlighting the results from the frequency count on high-quality and effective papers, BCTs identified in the other frequency counts are relevant because this presence is likely to reflect the feasibility, acceptability, or cost-effectiveness of a technique, aspects of an intervention that hold great importance.

These 7 BCTs should be focused on when designing web-based interventions in these addictive areas. It can be presumed that these BCTs were included because of their relevance to addictive behavior change, rather than purely from theoretical findings, as many interventions do not prioritize this theory and follow a systematic procedure when designing programs [102]. For example, self-monitoring of behavior (BCT item 2.3) is an important step in interventions to raise a user's awareness of their behavior and was identified in 86% of eligible papers. Similarly, providing instructions on how to perform a behavior (BCT item 4.1) was present in 55% of eligible papers. This is something that should be included in interventions that aim to provoke behavior change [103]. This means that the findings of this paper will be easily transferrable to the practical development of interventions, prompting to prioritize areas that should already have been considered by developers rather than causing web-based interventions to be completely redeveloped.

These results are consistent with other findings. For example, Michie et al [50] highlighted the role of the BCT self-monitoring of both behavior and outcomes of behavior in a healthy eating intervention. Both self-monitoring of behavior and self-monitoring of behavior outcomes were identified in the frequency counts of this review, with self-monitoring of behavior present in all frequency counts and self-monitoring of outcomes of behavior present in 3 of the 5 counts. The methodology of this study differed, with a meta-regression conducted by Michie et al [50] to examine the effectiveness of specific BCTs. However, we did not conduct a meta-analysis due to the heterogeneity of studies, and the aim of this paper is to identify few BCTs to inform further development

interventions, rather than focusing primarily on examining effectiveness and grouping BCTs together.

In a content analysis of alcohol reduction apps, Crane et al [48] identified 12 promising BCTs for effective behavior change. Similar to this study, self-monitoring of behavior, feedback on behavior, information about consequences, and social comparison were identified by the authors as useful BCTs. Instruction on how to perform the behavior, self-monitoring of behavior, and problem solving, which were identified in this study, were not identified as promising by Crane et al [48]. A similar level of support was found for the review by Rodda et al [49], with feedback on behavior, social comparison, and self-monitoring of behavior and outcomes of behavior identified as useful BCTs for behavior change in both this study and the review by Rodda et al [49]. Other BCTs were identified by authors, which this study did not support, for example, exposure and behavior substitution. However, inconsistent findings are to be expected in BCT research, as these techniques are highly dependent on target behavior, which were not identical.

Limitations

The effectiveness of BCTs is dependent on the behavior targeted by an intervention. This means that the results are only generalizable to these specific behaviors rather than behavior change on a wider scale. Despite this, it is encouraging that there were clearly common BCTs across the 3 identified problem behaviors, which can provide useful guidance for future intervention development. In addition to the BCTs used and the intervention's topic, multiple other variables contribute to the behavior change success of an intervention. These include motivation, opportunity, and perceived capability to change, which simultaneously act on behavior change and can be referred to as the COM-B (Capability, Opportunity, Motivation–Behavior) model [102]. The Behavior Change Wheel combines these behavioral sources with intervention functions, including education, persuasion and environmental restructuring, and policy categories such as guidelines and regulations, to provide an in-depth framework of behavior change [102]. Future research should take into account the interactions between these motivational and circumstantial variables and the mechanisms of behavior change. Without this consideration of factors including help seeking and resource availability, it is an oversimplification of behavior change theory and may lead to considerable resources being wasted on poorly implemented interventions. Therefore, this review can only provide guidance for intervention frameworks, rather than recommending a concrete intervention structure.

Although it provides an effective overview of BCTs, details in BCT taxonomy v1 are sometimes lacking and relatively subjective [104]. For example, BCT item 2.2 refers to feedback on behavior. This includes all possible feedback, with no specification on feedback type or delivery. Similarly, BCT item 5.1 is coded for any information about health consequences. Research has shown that the extremity of information content impacts one's behavioral response [105,106], which in turn will determine an intervention's behavior change effectiveness. This idea can be applied to the majority of the 93 BCTs identified within the taxonomy. In addition, there are other taxonomies, such as the Self-Enactable Techniques [107], which identify other techniques that may be relevant in behavior change interventions and are not directly listed in BCT taxonomy. For example, distraction (item 63), reflecting on the ability to perform behavior (item 99), and normalized behavior (item 114).

Finally, the scope of this paper was limited by the evidence base, which is biased toward positive, statistically significant publications [108]. It is possible that there are studies on behavior change interventions that used these BCTs but were not published due to the intervention finding no significant results [109]. Gray literature was not included in this review. The inclusion of this information is currently under debate, with a lack of consensus on how to conduct a systematic review of BCTs most effectively [99]. The exclusion of gray literature may limit the scope of a review, whereas inclusion may be able to bridge the research gap from publication bias as well as the time delay between conducting and publishing a study. However, including gray literature in a systematic review can lead to difficulties in data extraction and synthesis due to the lack of publishing standards being followed by authors [109]. This means that researcher bias is likely within data interpretation, justifying our exclusion of these sources. Furthermore, this review is limited by the interventions that exist and what is achievable with technology at the current time. In the future, it is expected that interventions will have more advanced features, which may lead to different BCT recommendations.

Implications

These results can inform the development of new web-based addiction treatments, providing recommendations of BCTs shown to be effective in past research to form the core of alcohol, binge eating, and gambling programs. This paper also highlights the importance of using evidence-based theory while developing behavior change interventions, which many treatments lack. Theory-based interventions are not only more likely to result in effective behavior change [110] but also allow a richer evaluation of interventions, enabling one to identify the active components of an intervention [102]. Ultimately, the consideration of behavior change theory and the testing of BCT effectiveness during the early stages of design means that addiction-based interventions as well as interventions targeting wider behaviors will result in greater, long-lasting behavior change [104]. Highlighted in the Behavior Change Wheel [102], behavior change is the result of an interacting system of individual and societal factors. Future research must adopt the experimental medicine approach [41,42] and examine not only whether these BCTs are present in interventions but also if they are the active mechanism of behavior change.

Conclusions

Alcohol consumption, binge eating behavior, and gambling can all be classified as compulsive or impulsive due to the neural mechanisms of reward that they impact. Owing to this comorbidity, transdiagnostic interventions may be used as a potential treatment. This systematic review identified the BCTs present in 42 web-based interventions that targeted one of these 3 behaviors. The authors explored the commonalities between

BCTs identified, controlling studies for quality and effectiveness, which can inform the development of future transdiagnostic interventions.

Acknowledgments

GH's PhD is funded in part by Oh My Mood and the Economic and Social Research Council UK.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms and individual frequency counts.
[DOCX File , 22 KB - jmir_v23i2e22694_app1.docx]

Multimedia Appendix 2
Study characteristics for eligible studies that targeted alcohol consumption.
[DOCX File , 34 KB - jmir_v23i2e22694_app2.docx]

Multimedia Appendix 3
Study characteristics for eligible studies that targeted gambling.
[DOCX File , 18 KB - jmir_v23i2e22694_app3.docx]

Multimedia Appendix 4
Study characteristics for eligible studies that targeted binge eating.
[DOCX File , 19 KB - jmir_v23i2e22694_app4.docx]

References

1. Calado F, Griffiths MD. Problem gambling worldwide: an update and systematic review of empirical research (2000–2015). J Behav Addict 2016 Dec;5(4):592-613. [doi: 10.1556/2006.5.2016.073]
2. Becker A. Sociocultural influences on body image and eating disturbance. In: Brownell KD, Walsh BT, editors. Eating Disorders and Obesity: A Comprehensive Handbook. New York City, NY: Guilford Press; 2018.
3. Peacock A, Leung J, Larney S, Colledge S, Hickman M, Rehm J, et al. Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. Addiction 2018 Oct 04;113(10):1905-1926 [FREE Full text] [doi: 10.1111/add.14234] [Medline: 29749059]
4. Lang B, Rosenberg H. Public perceptions of behavioral and substance addictions. Psychol Addict Behav 2017 Feb;31(1):79-84. [doi: 10.1037/adb0000228] [Medline: 27893214]
5. Rehm J. The risks associated with alcohol use and alcoholism. Alcohol Res Health 2011;34(2):135-143 [FREE Full text] [Medline: 22330211]
6. The NHS long term plan. National Health Service. 2019 Jan. URL: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/01/nhs-long-term-plan-june-2019.pdf [accessed 2021-01-13]
7. WHO. Global status report on alcohol and health 2014. World Health Organization. 2014. URL: https://www.who.int/substance_abuse/publications/global_alcohol_report/msb_gsr_2014_1.pdf?ua=1 [accessed 2021-01-13]
8. Sharma A, Sinha K, Vandenberg B. Pricing as a means of controlling alcohol consumption. Br Med Bull 2017 Sep 01;123(1):149-158. [doi: 10.1093/bmb/ldx020] [Medline: 28910991]
9. Drinking behaviours among adults. Statistics on Alcohol, England 2020. Leeds: NHS Digital URL: https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-alcohol/2020/part-4 [accessed 2021-01-13]
10. World Health Organization. Global status report on alcohol and health 2018. Geneva: WHO; 2018:1-450.
11. Panayotova P. Beyond drink, drainage and divorce? What the Sociological Society did for British sociology. The Sociological Review 2019 Apr 29;67(6):1212-1227. [doi: 10.1177/0038026119846783]
12. Cinquegrani C, Brown DHK. 'Wellness' lifts us above the Food Chaos': a narrative exploration of the experiences and conceptualisations of Orthorexia Nervosa through online social media forums. Qualitative Research in Sport, Exercise and Health 2018 Apr 27;10(5):585-603. [doi: 10.1080/2159676x.2018.1464501]
13. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5). American Psychiatric Publishing 2013. [doi: 10.1176/appi.books.9780890425596]
14. Bongers P, Jansen A. Emotional eating and Pavlovian learning: evidence for conditioned appetitive responding to negative emotional states. Cogn Emot 2017 Feb 05;31(2):284-297. [doi: 10.1080/02699931.2015.1108903] [Medline: 26539994]

15. Mental Health Bulletin 2018-19 Annual Report. NHS Digital. Leeds: NHS Digital; 2019 Nov 28. URL: https://files.digital.nhs.uk/98/150C99/MHB-1819-Annual%20Report.pdf [accessed 2021-01-13]
16. Eating disorders in young people. Mental health of children and young people. London: Public Health England; 2019 Jul 12. URL: https://www.gov.uk/government/publications/eating-disorders-in-young-people [accessed 2021-01-13]
17. Šmahel D, Macháčková H, Šmahelová M, Čevelíček M, Almenara CA, Holubčíková J. What healthcare professionals should know about digital technologies and eating disorders. In: Digital Technology, Eating Behaviors, and Eating Disorders. Switzerland: Springer, Cham; 2018:139-157.
18. Saunders JF, Eaton AA. Snaps, selfies, and shares: how three popular social media platforms contribute to the sociocultural model of disordered eating among young women. Cyberpsychol Behav Soc Netw 2018 Jun;21(6):343-354. [doi: 10.1089/cyber.2017.0713] [Medline: 29883209]
19. Kircaburun K, Yurdagül C, Kuss D, Emirtekin E, Griffiths MD. Problematic mukbang watching and its relationship to disordered eating and internet addiction: a pilot study among emerging adult mukbang watchers. Int J Ment Health Addiction 2020 May 21:-. [doi: 10.1007/s11469-020-00309-w]
20. Chóliz M. The challenge of online gambling: the effect of legalization on the increase in online gambling addiction. J Gambl Stud 2015 Jun 10;32(2):749-756. [doi: 10.1007/s10899-015-9558-6]
21. Essau CA. Comorbidity of addictive problems: assessment and treatment implications. In: Adolescent Addiction (Second Edition). Switzerland: Elsevier B.V; 2020:291-317.
22. Reyna VF, Huettel SA. Reward, representation, and impulsivity: a theoretical framework for the neuroscience of risky decision making. American Psychological Association 2014:11-42. [doi: 10.1037/14322-002]
23. Czernecka K, Fryt J, Szczygieł M, Torre AL, Smolen T. Many face(t)s of young people's risk-taking: individual and situational determinants. CIPP 2015:-. [doi: 10.5114/cipp.2018.72268]
24. Fauth-Bühler M. Similarities and differences between gambling disorder and other addiction-like behaviors. In Gambling Disorder 2019;235:246. [doi: 10.1007/978-3-030-03060-5_11]
25. Fusar-Poli P, Solmi M, Brondino N, Davies C, Chae C, Politi P, et al. Transdiagnostic psychiatry: a systematic review. World Psychiatry 2019 Jun 06;18(2):192-207 [FREE Full text] [doi: 10.1002/wps.20631] [Medline: 31059629]
26. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev 2015 Aug;40:91-110 [FREE Full text] [doi: 10.1016/j.cpr.2015.06.002] [Medline: 26094079]
27. Robinson SM. "Alcoholic" or "Person with alcohol use disorder"? Applying person-first diagnostic terminology in the clinical domain. Subst Abus 2017;38(1):9-14. [doi: 10.1080/08897077.2016.1268239] [Medline: 27925860]
28. Griffiths S, Mond JM, Li Z, Gunatilake S, Murray SB, Sheffield J, et al. Self-stigma of seeking treatment and being male predict an increased likelihood of having an undiagnosed eating disorder. Int J Eat Disord 2015 Sep 06;48(6):775-778. [doi: 10.1002/eat.22413] [Medline: 26052695]
29. Dąbrowska K, Wieczorek L. Perceived social stigmatisation of gambling disorders and coping with stigma. Nordic Studies on Alcohol and Drugs 2020 Mar 01;37(3):279-297. [doi: 10.1177/1455072520902342]
30. Livingstone C, Adams P, Cassidy R, Markham F, Reith G, Rintoul A, et al. On gambling research, social science and the consequences of commercial gambling. International Gambling Studies 2017 Sep 27;18(1):56-68. [doi: 10.1080/14459795.2017.1377748]
31. Kärkkäinen U, Mustelin L, Raevuori A, Kaprio J, Keski-Rahkonen A. Do disordered eating behaviours have long-term health-related consequences? Eur Eat Disord Rev 2018 Jan 21;26(1):22-28 [FREE Full text] [doi: 10.1002/erv.2568] [Medline: 29160017]
32. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
33. Kazdin AE. Addressing the treatment gap: a key challenge for extending evidence-based psychosocial interventions. Behav Res Ther 2017 Jan;88:7-18. [doi: 10.1016/j.brat.2016.06.004] [Medline: 28110678]
34. Albright L. A novel three-tiered approach to bridging the mental health treatment gap. Population Health Research Award for Undergraduates 2019:-.
35. Cohen J. Statistical Power Analysis. Curr Dir Psychol Sci 2016 Jun 24;1(3):98-101. [doi: 10.1111/1467-8721.ep10768783]
36. Godfrey KM, Gallo LC, Afari N. Mindfulness-based interventions for binge eating: a systematic review and meta-analysis. J Behav Med 2015 Apr 23;38(2):348-362. [doi: 10.1007/s10865-014-9610-5] [Medline: 25417199]
37. Carey K, Scott-Sheldon L, Elliott J, Bolles J, Carey M. Computer-delivered interventions to reduce college student drinking: a meta-analysis. Addiction 2009 Nov;104(11):1807-1819 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02691.x] [Medline: 19744139]
38. Gooding P, Tarrier N. A systematic review and meta-analysis of cognitive-behavioural interventions to reduce problem gambling: hedging our bets? Behav Res Ther 2009 Jul;47(7):592-607. [doi: 10.1016/j.brat.2009.04.002] [Medline: 19446287]
39. Feeley N, Cossette S, Côté J, Héon M, Stremler R, Martorella G, et al. The importance of piloting an RCT intervention. Can J Nurs Res 2009 Jun;41(2):85-99. [Medline: 19650515]

40. Tebb KP, Erenrich RK, Jasik CB, Berna MS, Lester JC, Ozer EM. Use of theory in computer-based interventions to reduce alcohol use among adolescents and young adults: a systematic review. BMC Public Health 2016 Jun 17;16(1):517 [FREE Full text] [doi: 10.1186/s12889-016-3183-x] [Medline: 27317330]
41. Sheeran P, Klein WM, Rothman AJ. Health behavior change: moving from observation to intervention. Annu Rev Psychol 2017 Jan 03;68(1):573-600. [doi: 10.1146/annurev-psych-010416-044007] [Medline: 27618942]
42. Field M, Christiansen P, Hardman C, Haynes A, Jones A, Robinson E, et al. Translation of laboratory research into behavioral interventions for obesity and problem drinking: The experimental medicine approach. Researchgate 2019 Sep:- [FREE Full text] [doi: 10.31234/osf.io/scre4]
43. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
44. Chen J, Liu X, Rapee RM, Pillay P. Behavioural activation: a pilot trial of transdiagnostic treatment for excessive worry. Behav Res Ther 2013 Sep;51(9):533-539. [doi: 10.1016/j.brat.2013.05.010] [Medline: 23827657]
45. Lorencatto F, West R, Stavri Z, Michie S. How well is intervention content described in published reports of smoking cessation interventions? Nicotine Tob Res 2013 Jul 21;15(7):1273-1282. [doi: 10.1093/ntr/nts266] [Medline: 23262584]
46. Michie S, Abraham C. Interventions to change health behaviours: evidence-based or evidence-inspired? Psychology & Health 2004 Feb;19(1):29-49. [doi: 10.1080/0887044031000141199]
47. Prestwich A, Sniehotta FF, Whittington C, Dombrowski SU, Rogers L, Michie S. Does theory influence the effectiveness of health behavior interventions? Meta-analysis. Health Psychol 2014 May;33(5):465-474. [doi: 10.1037/a0032853] [Medline: 23730717]
48. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015 May 14;17(5):118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
49. Rodda S, Merkouris SS, Abraham C, Hodgins DC, Cowlishaw S, Dowling NA. Therapist-delivered and self-help interventions for gambling problems: a review of contents. J Behav Addict 2018 Jun 01;7(2):211-226 [FREE Full text] [doi: 10.1556/2006.7.2018.44] [Medline: 29895185]
50. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
51. A systematic review examining effective behaviour change techniques in alcohol, emotional eating and gambling online interventions to inform the development of transdiagnostic programmes. Center for Open Science. URL: https://osf.io/uxe67/ [accessed 2021-01-25]
52. Bewick BM, West R, Gill J, O'May F, Mulhern B, Barkham M, et al. Providing web-based feedback and social norms information to reduce student alcohol intake: a multisite investigation. J Med Internet Res 2010 Dec 19;12(5):59 [FREE Full text] [doi: 10.2196/jmir.1461] [Medline: 21169171]
53. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
54. Sobell LC, Sobell MB. Timeline follow-back. Measuring alcohol consumption 1992;41:72. [doi: 10.1007/978-1-4612-0357-5_3]
55. Crombie IK, Irvine L, Williams B, Sniehotta FF, Petrie D, Jones C, et al. Texting to Reduce Alcohol Misuse (TRAM): main findings from a randomized controlled trial of a text message intervention to reduce binge drinking among disadvantaged men. Addiction 2018 Jun 01;113(9):1609-1618. [doi: 10.1111/add.14229] [Medline: 29855105]
56. Pedersen ER, Parast L, Marshall GN, Schell TL, Neighbors C. A randomized controlled trial of a web-based, personalized normative feedback alcohol intervention for young-adult veterans. J Consult Clin Psychol 2017 May;85(5):459-470 [FREE Full text] [doi: 10.1037/ccp0000187] [Medline: 28287799]
57. Andrade ALM, de Lacerda RB, Gomide HP, Ronzani TM, Sartes LMA, Martins LF, WHO e-health project on AlcoholHealth Investigators Group. Web-based self-help intervention reduces alcohol consumption in both heavy-drinking and dependent alcohol users: a pilot study. Addict Behav 2016 Dec;63:63-71. [doi: 10.1016/j.addbeh.2016.06.027] [Medline: 27424165]
58. Arnaud N, Baldus C, Elgán TH, De Paepe N, Tønnesen H, Csémy L, et al. Effectiveness of a web-based screening and fully automated brief motivational intervention for adolescent substance use: a randomized controlled trial. J Med Internet Res 2016 May 24;18(5):103 [FREE Full text] [doi: 10.2196/jmir.4643] [Medline: 27220276]
59. Bertholet N, Cunningham JA, Faouzi M, Gaume J, Gmel G, Burnand B, et al. Internet-based brief intervention for young men with unhealthy alcohol use: a randomized controlled trial in a general population sample. Addiction 2015 Nov;110(11):1735-1743. [doi: 10.1111/add.13051] [Medline: 26173842]
60. Bertholet N, Daeppen J, McNeely J, Kushnir V, Cunningham JA. Smartphone application for unhealthy alcohol use: a pilot study. Subst Abus 2017 Jan 23;38(3):285-291. [doi: 10.1080/08897077.2017.1281860] [Medline: 28113039]
61. Bewick BM, West RM, Barkham M, Mulhern B, Marlow R, Traviss G, et al. The effectiveness of a web-based personalized feedback and social norms alcohol intervention on United Kingdom university students: randomized controlled trial. J Med Internet Res 2013 Jul 24;15(7):137 [FREE Full text] [doi: 10.2196/jmir.2581] [Medline: 23883616]

62. Bingham CR, Barretto AI, Walton MA, Bryant CM, Shope JT, Raghunathan TE. Efficacy of a web-based, tailored, alcohol prevention/intervention program for college students: 3-month follow-up. J Drug Educ 2012 Feb 06;41(4):405-430. [doi: 10.2190/de.41.4.e]
63. Deady M, Mills KL, Teesson M, Kay-Lambkin F. An online intervention for co-occurring depression and problematic alcohol use in young people: primary outcomes from a randomized controlled trial. J Med Internet Res 2016 Mar 23;18(3):71 [FREE Full text] [doi: 10.2196/jmir.5178] [Medline: 27009465]
64. Doumas D, Haustveit T. Reducing heavy drinking in intercollegiate athletes valuation of a web-based personalized feedback program. Sport Psychol 2008;22(2):-. [doi: 10.1123/tsp.22.2.212]
65. Doumas D, Kane C, Navarro T, Roman J. Decreasing heavy drinking in first-year students valuation of a web-based personalized feedback program administered during orientation. Journal of College Counseling 2011;14(1):-. [doi: 10.1002/j.2161-1882.2011.tb00060.x]
66. Dulin PL, Gonzalez VM, Campbell K. Results of a pilot test of a self-administered smartphone-based treatment system for alcohol use disorders: usability and early outcomes. Subst Abus 2014;35(2):168-175 [FREE Full text] [doi: 10.1080/08897077.2013.821437] [Medline: 24821354]
67. Fazzino TL, Rose GL, Helzer JE. An experimental test of assessment reactivity within a web-based brief alcohol intervention study for college students. Addict Behav 2016 Jan;52:66-74 [FREE Full text] [doi: 10.1016/j.addbeh.2015.08.011] [Medline: 26363306]
68. Ganz T, Braun M, Laging M, Schermelleh-Engel K, Michalak J, Heidenreich T. Effects of a stand-alone web-based electronic screening and brief intervention targeting alcohol use in university students of legal drinking age: a randomized controlled trial. Addict Behav 2018 Feb;77:81-88. [doi: 10.1016/j.addbeh.2017.09.017] [Medline: 28985586]
69. Haug S, Schaub MP, Venzin V, Meyer C, John U, Gmel G. A pre-post study on the appropriateness and effectiveness of a web- and text messaging-based intervention to reduce problem drinking in emerging adults. J Med Internet Res 2013 Sep 02;15(9):196 [FREE Full text] [doi: 10.2196/jmir.2755] [Medline: 23999406]
70. Jander A, Crutzen R, Mercken L, Candel M, de Vries H. Effects of a web-based computer-tailored game to reduce binge drinking among dutch adolescents: a cluster randomized controlled trial. J Med Internet Res 2016 Feb 03;18(2):29 [FREE Full text] [doi: 10.2196/jmir.4708] [Medline: 26842694]
71. Jo S, Lee HK, Kang K, Joe KH, Lee S. Efficacy of a web-based screening and brief intervention to prevent problematic alcohol use in korea: results of a randomized controlled trial. Alcohol Clin Exp Res 2019 Oct;43(10):2196-2202. [doi: 10.1111/acer.14169] [Medline: 31386203]
72. Kypri K, Langley JD, Saunders JB, Cashell-Smith ML, Herbison P. Randomized controlled trial of web-based alcohol screening and brief intervention in primary care. Arch Intern Med 2008 Mar 10;168(5):530-536. [doi: 10.1001/archinternmed.2007.109] [Medline: 18332300]
73. Kypri K, Hallett J, Howat P, McManus A, Maycock B, Bowe S, et al. Randomized controlled trial of proactive web-based alcohol screening and brief intervention for university students. Arch Intern Med 2009 Sep 14;169(16):1508-1514. [doi: 10.1001/archinternmed.2009.249] [Medline: 19752409]
74. Kypri K, McCambridge J, Vater T, Bowe SJ, Saunders JB, Cunningham JA, et al. Web-based alcohol intervention for Māori university students: double-blind, multi-site randomized controlled trial. Addiction 2013 Feb;108(2):331-338 [FREE Full text] [doi: 10.1111/j.1360-0443.2012.04067.x] [Medline: 22925046]
75. Palfai TP, Winter M, Lu J, Rosenbloom D, Saitz R. Personalized feedback as a universal prevention approach for college drinking: a randomized trial of an e-mail linked universal web-based alcohol intervention. J Prim Prev 2014 Apr 14;35(2):75-84 [FREE Full text] [doi: 10.1007/s10935-013-0337-9] [Medline: 24421075]
76. Paschall MJ, Antin T, Ringwalt CL, Saltz RF. Evaluation of an Internet-based alcohol misuse prevention course for college freshmen: findings of a randomized multi-campus trial. Am J Prev Med 2011 Sep;41(3):300-308 [FREE Full text] [doi: 10.1016/j.amepre.2011.03.021] [Medline: 21855745]
77. Possemato K, Johnson EM, Emery JB, Wade M, Acosta MC, Marsch LA, et al. A pilot study comparing peer supported web-based CBT to self-managed web CBT for primary care veterans with PTSD and hazardous alcohol use. Psychiatr Rehabil J 2019 Sep;42(3):305-313 [FREE Full text] [doi: 10.1037/prj0000334] [Medline: 30489140]
78. Schulz DN, Candel MJ, Kremers SP, Reinwand DA, Jander A, de Vries H. Effects of a Web-based tailored intervention to reduce alcohol consumption in adults: randomized controlled trial. J Med Internet Res 2013 Sep 17;15(9):206 [FREE Full text] [doi: 10.2196/jmir.2568] [Medline: 24045005]
79. Sinadinovic K, Wennberg P, Johansson M, Berman AH. Targeting individuals with problematic alcohol use via web-based cognitive-behavioral self-help modules, personalized screening feedback or assessment only: a randomized controlled trial. Eur Addict Res 2014 Oct 4;20(6):305-318 [FREE Full text] [doi: 10.1159/000362406] [Medline: 25300885]
80. Spijkerman R, Roek MAE, Vermulst A, Lemmers L, Huiberts A, Engels RCME. Effectiveness of a web-based brief alcohol intervention and added value of normative feedback in reducing underage drinking: a randomized controlled trial. J Med Internet Res 2010 Dec 19;12(5):65 [FREE Full text] [doi: 10.2196/jmir.1465] [Medline: 21169172]
81. Suffoletto B, Kristan J, Callaway C, Kim KH, Chung T, Monti PM, et al. A text message alcohol intervention for young adult emergency department patients: a randomized clinical trial. Ann Emerg Med 2014 Dec;64(6):664-724 [FREE Full text] [doi: 10.1016/j.annemergmed.2014.06.010] [Medline: 25017822]

82. Voogt CV, Kuntsche E, Kleinjan M, Poelen EAP, Lemmers LACJ, Engels RCME. Using ecological momentary assessment in testing the effectiveness of an alcohol intervention: a two-arm parallel group randomized controlled trial. PLoS One 2013 Nov 5;8(11):78436 [FREE Full text] [doi: 10.1371/journal.pone.0078436] [Medline: 24223806]
83. Voogt CV, Poelen EAP, Kleinjan M, Lemmers LACJ, Engels RCME. The effectiveness of the 'what do you drink' web-based brief alcohol intervention in reducing heavy drinking among students: a two-arm parallel group randomized controlled trial. Alcohol Alcohol 2013;48(3):312-321. [doi: 10.1093/alcalc/ags133] [Medline: 23303466]
84. Wallace P, Murray E, McCambridge J, Khadjesari Z, White IR, Thompson SG, et al. On-line randomized controlled trial of an internet based psychologically enhanced intervention for people with hazardous alcohol consumption. PLoS One 2011 Mar 09;6(3):14740 [FREE Full text] [doi: 10.1371/journal.pone.0014740] [Medline: 21408060]
85. Wilks C, Lungu A, Ang S, Matsumiya B, Yin Q, Linehan M. A randomized controlled trial of an internet delivered dialectical behavior therapy skills training for suicidal and heavy episodic drinkers. J Affect Disord 2018;232:219-228. [doi: 10.2196/preprints.7767]
86. Bücker L, Bierbrodt J, Hand I, Wittekind C, Moritz S. Effects of a depression-focused internet intervention in slot machine gamblers: a randomized controlled trial. PLoS ONE 2018 Jun 8;13(6):0198859. [doi: 10.1371/journal.pone.0198859]
87. Canale N, Vieno A, Griffiths MD, Marino C, Chieco F, Disperati F, et al. The efficacy of a web-based gambling intervention program for high school students: a preliminary randomized study. Computers in Human Behavior 2016 Feb;55:946-954. [doi: 10.1016/j.chb.2015.10.012]
88. Casey LM, Oei TPS, Raylu N, Horrigan K, Day J, Ireland M, et al. Internet-based delivery of cognitive behaviour therapy compared to monitoring, feedback and support for problem gambling: a randomised controlled trial. J Gambl Stud 2017 Sep 25;33(3):993-1010. [doi: 10.1007/s10899-016-9666-y] [Medline: 28124288]
89. Hodgins DC, Cunningham JA, Murray R, Hagopian S. Online self-directed interventions for gambling disorder: randomized controlled trial. J Gambl Stud 2019 Jun 30;35(2):635-651. [doi: 10.1007/s10899-019-09830-7] [Medline: 30701377]
90. Ivanova E, Magnusson K, Carlbring P. Deposit limit prompt in online gambling for reducing gambling intensity: a randomized controlled trial. Front Psychol 2019 Mar 28;10:639 [FREE Full text] [doi: 10.3389/fpsyg.2019.00639] [Medline: 31001160]
91. Neighbors C, Rodriguez LM, Rinker DV, Gonzales RG, Agana M, Tackett JL, et al. Efficacy of personalized normative feedback as a brief intervention for college student gambling: a randomized controlled trial. J Consult Clin Psychol 2015 Jun;83(3):500-511 [FREE Full text] [doi: 10.1037/a0039125] [Medline: 26009785]
92. Wood RT, Wohl MJ. Assessing the effectiveness of a responsible gambling behavioural feedback tool for reducing the gambling expenditure of at-risk players. International Gambling Studies 2015 Jul 29;15(2):1-16. [doi: 10.1080/14459795.2015.1049191]
93. Boucher S, Edwards O, Gray A, Nada-Raja S, Lillis J, Tylka TL, et al. Teaching intuitive eating and acceptance and commitment therapy skills via a web-based intervention: a pilot single-arm intervention study. JMIR Res Protoc 2016 Oct 14;5(4):180 [FREE Full text] [doi: 10.2196/resprot.5861] [Medline: 27742602]
94. Carrard I, Crépin C, Rouget P, Lam T, Golay A, Van der Linden M. Randomised controlled trial of a guided self-help treatment on the Internet for binge eating disorder. Behav Res Ther 2011 Aug;49(8):482-491. [doi: 10.1016/j.brat.2011.05.004] [Medline: 21641580]
95. Lyzwinski LN, Caffery L, Bambling M, Edirippulige S. The mindfulness app trial for weight, weight-related behaviors, and stress in university students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Apr 10;7(4):12210 [FREE Full text] [doi: 10.2196/12210] [Medline: 30969174]
96. Ruwaard J, Lange A, Broeksteeg J, Renteria-Agirre A, Schrieken B, Dolan CV, et al. Online cognitive-behavioural treatment of bulimic symptoms: a randomized controlled trial. Clin Psychol Psychother 2013;20(4):308-318. [doi: 10.1002/cpp.1767] [Medline: 22298417]
97. Svensson M, Hult M, van der Mark M, Grotta A, Jonasson J, von Hausswolff-Juhlin Y, et al. The change in eating behaviors in a Web-based weight loss program: a longitudinal analysis of study completers. J Med Internet Res 2014 Nov 03;16(11):234 [FREE Full text] [doi: 10.2196/jmir.3131] [Medline: 25367316]
98. Völker U, Jacobi C, Taylor CB. Adaptation and evaluation of an internet-based prevention program for eating disorders in a sample of women with subclinical eating disorder symptoms: a pilot study. Eating Weight Disord 2014 Apr 23;16(4):270-273. [doi: 10.1007/bf03327471]
99. Armitage CJ, Conner M, Prestwich A, de Bruin M, Johnston M, Sniehotta F, et al. Investigating which behaviour change techniques work for whom in which contexts delivered by what means: proposal for an international collaboratory of Centres for Understanding Behaviour Change (CUBiC). Br J Health Psychol 2020 Oct 20:-. [doi: 10.1111/bjhp.12479] [Medline: 33080120]
100. Boot WR, Simons DJ, Stothart C, Stutts C. The pervasive problem with placebos in psychology: why active control groups are not sufficient to rule out placebo effects. Perspect Psychol Sci 2013 Jul;8(4):445-454. [doi: 10.1177/1745691613491271] [Medline: 26173122]
101. Office of Health Assessment and Translation. OHAT risk of bias rating tool for human and animal studies. National Toxicology Program 2015:1-37 [FREE Full text]
102. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]

103. Presseau J, Ivers NM, Newham JJ, Knittle K, Danko KJ, Grimshaw JM. Using a behaviour change techniques taxonomy to identify active ingredients within trials of implementation interventions for diabetes care. Implement Sci 2015 Apr 23;10(1):55 [FREE Full text] [doi: 10.1186/s13012-015-0248-7] [Medline: 25900104]
104. Lorencatto F, Smith L, Hamer M, Biddle S. Applying the BCT taxonomy to code sedentary behaviour reduction interventions: challenges and reflections. European Health Psychologist 2016;18:442 [FREE Full text]
105. Boshoff C, Toerien L. Subconscious responses to fear-appeal health warnings: an exploratory study of cigarette packaging. SAJEMS 2017 Apr 28;20(1). [doi: 10.4102/sajems.v20i1.1630]
106. Ferestad J, Thompson M. "I'm Not Gonna Be Like 'That Guy'": examining antidrug shock advertising through the eyes of "That Guy". Deviant Behavior 2016 Jul 01;38(2):173-187. [doi: 10.1080/01639625.2016.1196971]
107. Knittle K, Heino M, Marques MM, Stenius M, Beattie M, Ehbrecht F, et al. The compendium of self-enactable techniques to change and self-manage motivation and behaviour v.1.0. Nat Hum Behav 2020 Feb;4(2):215-223. [doi: 10.1038/s41562-019-0798-9] [Medline: 31932687]
108. Pappas C, Williams I. Grey literature: its emerging importance. J of Hosp Librarianship 2011 Jul;11(3):228-234. [doi: 10.1080/15323269.2011.587100]
109. Adams J, Hillier-Brown FC, Moore HJ, Lake AA, Araujo-Soares V, White M, et al. Searching and synthesising 'grey literature' and 'grey information' in public health: critical reflections on three case studies. Syst Rev 2016 Sep 29;5(1):164 [FREE Full text] [doi: 10.1186/s13643-016-0337-y] [Medline: 27686611]
110. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]

Abbreviations

BCT: behavior change technique
OHAT: Office of Health Assessment and Translation
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 20.07.20; peer-reviewed by S Rodda, J Ayre; comments to author 19.10.20; revised version received 01.12.20; accepted 07.12.20; published 09.02.21.

Please cite as:
Humphreys G, Evans R, Makin H, Cooke R, Jones A
Identification of Behavior Change Techniques From Successful Web-Based Interventions Targeting Alcohol Consumption, Binge Eating, and Gambling: Systematic Review
J Med Internet Res 2021;23(2):e22694
URL: http://www.jmir.org/2021/2/e22694/
doi:10.2196/22694
PMID:33560243

Original Paper

Recommendations for Designing Health Information Technologies for Mental Health Drawn From Self-Determination Theory and Co-design With Culturally Diverse Populations: Template Analysis

Vanessa Wan Sze Cheng[1], PhD; Sarah E Piper[1], GradDip, Psych(Hons); Antonia Ottavio[1], DipHthSc, BN; Tracey A Davenport[1], BA(Hons), eMBA; Ian B Hickie[1], AM, MD, FRANZCP, FASSA

Brain and Mind Centre, The University of Sydney, Sydney, Australia

Corresponding Author:
Vanessa Wan Sze Cheng, PhD
Brain and Mind Centre
The University of Sydney
94 Mallett Street
Camperdown
Sydney, 2050
Australia
Phone: 61 93510774
Email: vanessa.cheng@sydney.edu.au

Abstract

Background: Culturally diverse populations (including Aboriginal and Torres Strait Islander people, people of diverse genders and sexualities, and culturally and linguistically diverse people) in nonurban areas face compounded barriers to accessing mental health care. Health information technologies (HITs) show promising potential to overcome these barriers.

Objective: This study aims to identify how best to improve a mental health and well-being HIT for culturally diverse Australians in nonurban areas.

Methods: We conducted 10 co-design workshops (N=105 participants) in primary youth mental health services across predominantly nonurban areas of Australia and conducted template analysis on the workshop outputs. Owing to local (including service) demographics, the workshop participants naturalistically reflected culturally diverse groups.

Results: We identified 4 main themes: control, usability, affirmation, and health service delivery factors. The first 3 themes overlap with the 3 basic needs postulated by self-determination theory (autonomy, competence, and relatedness) and describe participant recommendations on how to *design* an HIT. The final theme includes barriers to adopting HITs for mental health care and how HITs can be used to support care coordination and delivery. Hence, it describes participant recommendations on how to *use* an HIT.

Conclusions: Although culturally diverse groups have specific concerns, their expressed needs fall broadly within the relatively universal design principles identified in this study. The findings of this study provide further support for applying self-determination theory to the design of HITs and reflect the tension in designing technologies for complex problems that overlap multiple medical, regulatory, and social domains, such as mental health care. Finally, we synthesize the identified themes into general recommendations for designing HITs for mental health and provide concrete examples of design features recommended by participants.

(J Med Internet Res 2021;23(2):e23502) doi:10.2196/23502

KEYWORDS

mental health; health information technologies; self-determination theory; eHealth; internet; digital health; adolescent; mental health services; young adult; LGBTQ persons; mobile phone; rural health

Introduction

Barriers to Accessing Mental Health Care for Young Nonurban Australians

Around 1 in 5 Australians experience mental ill health regardless of whether they live in an urban or a nonurban (ie, regional, rural, or remote) area [1]. However, compared with urban Australians, nonurban Australians experience worse mental health outcomes, such as higher self-harm, suicidal ideation, and suicide attempt rates [2]. Nonurban Australians experience many barriers to accessing services and supports, such as cost, distance, a relative lack of service providers, stigma, and an increased emphasis on privacy [3,4].

These barriers can lead to limited access to mental health care for young people living in nonurban areas. Notably, rural young Australians discontinue prematurely from care at higher rates than those from urban or regional areas [5]. As service provision declines with distance from major cities [6], the lack of public transportation in rural and remote communities increases young people's reliance on their supportive others (eg, parents) to take them to a service [7]. This reliance on others to access a service also raises the issue of anonymity, which can be problematic given the higher levels of stigma related to mental health issues in nonurban communities [4]. Furthermore, service opening times (often limited to weekdays during 9 AM to 5 PM), extensive wait times, and cost can add to these barriers to access mental health care [8].

Culturally Diverse Young People

For culturally diverse young nonurban Australians, such as Aboriginal and Torres Strait Islander people, people of diverse genders and sexualities, inclusive of and not limited to lesbian, gay, bisexual, transgender, queer, intersex, asexual, questioning, and pansexual people (henceforth referred to as LGBTQIA+ people in this paper), and culturally and linguistically diverse (CALD) people, multiple forms of inequality often intersect to create compounded barriers in the form of decreased mental health literacy, financial barriers, increased social and self-stigma, and a lack of mental health services compounded with geographic inaccessibility [7]. These culturally diverse groups experience both poorer mental health outcomes and reduced access to mental health care [9-13].

Barriers to accessing mental health care for Aboriginal and Torres Strait Islander people living in nonurban areas include a lack of trust in health services, lack of culturally appropriate care, and lack of available services in remote areas [10,14-16]. For young LGBTQIA+ people, reported barriers include concerns that health professionals would not be able to cater to an individual's specific identity or needs [17], a fear of experiencing homophobia and/or transphobia, and a reluctance to come out to a health professional [18]. Finally, it has been reported that young CALD people, including young people from migrant and refugee backgrounds, underutilize the public mental health system in Australia [19,20]. The key barriers to mental health care access for young CALD people include the stigma associated with mental illness, concerns regarding confidentiality, limited knowledge of available services, language barriers and communication difficulties, fear of discrimination, and a lack of trust in service providers [19,21]. These barriers to appropriate care for these 3 populations can be exacerbated in a rural setting because of fewer available culturally competent services and the increased emphasis on privacy within close-knit rural communities [4,7].

Technologies to Empower Mental Health Care Access

The term health information technology (HIT) has been defined as the "application of information processing involving both computer hardware and software that deals with the storage, retrieval, sharing, and use of health care information, data, and knowledge for communication and decision making" [22]. HITs have immense potential to address intersecting barriers to mental health help seeking in nonurban areas. They can provide individuals with access to mental health services regardless of geographical location, vulnerability, or socioeconomic status [23]; reduce the wait time to access mental health support; and remove constraints around service opening hours [24]. HITs can also make psychoeducation and help-seeking sources easier to locate and access [23]. Not only can HITs enable access to care, but they can also enhance the care delivered by mental health services, for example by using synchronous communication protocols such as video chat and real-time data tracking through apps and other software. However, if HITs do not achieve a minimum standard of usability (or ease of use) as well as user experience, they experience suboptimal uptake and dropout rates [23,25]. Given the varying levels of satisfaction (or lack thereof) of young people toward web-based mental health resources [26], it is important to ensure that HITs support the needs of their users appropriately.

To investigate how HITs can improve and transform mental health care in Australia, the University of Sydney's Brain and Mind Centre (BMC) established the Youth Mental Health and Technology Program (YMH and Tech Program). As part of this program, an HIT prototype known as the BMC Youth Platform was developed to enhance the quality of health care provided by traditional mental health services. The BMC Youth Platform consists of a set of web-based personalized clinical assessments and longitudinal tracking tools for young people to monitor psychological, neurocognitive, social, and medical characteristics and plan individualized and more effective longer-term interventions. It aims to support the delivery of mental health services and the management of mental health symptoms and as such is to be used by young people, their health professionals, and their supportive others. A summary of the key BMC Youth Platform components is presented in Table 1.

Table 1. Key components of the Brain and Mind Youth Platform.

Purpose	Component
Start page	Start page
Record of basic demographic information	About me page
Record of medical history	Health history page
Clinical assessment	Questionnaires
Holistic multidimensional tracking of mental health	Dashboard of multiple health cards (each representing a different health domain)
Recommended interventions	Care options

The BMC Youth Platform was implemented into 5 *headspace* centers (a primary mental health service for young people aged between 12 and 25 years) in urban locations within Sydney, New South Wales, and learnings from these implementations included the importance of ongoing co-design with end users to ensure iterative improvements to the HIT [24]. Our past research has also confirmed the suitability of, and the potential to further explore, the use of co-design methodologies when implementing HITs in nonurban areas of Australia. Crucially, co-design methodologies have been found to facilitate the implementation of more acceptable digital solutions for mental health in nonurban areas that take local circumstances (such as natural disaster rates) into account [27].

Iteratively Co-designing and Testing HITs

The BMC Youth Platform is the product of multiple years of co-design and user testing conducted, as part of the YMH and Tech Program, with various representative end user groups, including young people aged between 16 and 25 years and health professionals [28,29]. Co-design (also referred to as participatory design) is a key research methodology that enables the perspectives and preferences of the target end user population to influence subsequent development of the HIT [30]. When conducted appropriately, participatory design is effective in obtaining insights from population groups that are marginalized or otherwise affected by structural inequalities [31,32] and results in higher levels of end user acceptability of the final intervention [33].

One product of the YMH and Tech Program is the Project Synergy Research and Development Cycle, which applies co-design methodologies to the design, development, implementation, and feasibility testing of apps and technologies [28]. In total, 3 key principles underpin this cycle: involving target end user populations (including, but not limited to, young people, supportive others, health professionals, and other service staff) as active participants throughout the entire design process, treating young people as design partners, and continually and iteratively evaluating the acceptability of the technology from the perspective of its target audience.

Although previous co-design and user testing work done as part of the YMH and Tech Program focused on the needs of young Australians in urban areas [28] and expanded them to young Australians in nonurban areas [27], these methodologies have not yet been applied to culturally diverse young Australians in nonurban areas.

Objectives

This study aims to identify how best to improve an HIT (such as the BMC Youth Platform) for culturally diverse young Australians in nonurban areas and to synthesize findings into recommendations for designing HITs for mental health.

Methods

Ethical Approval

The University of Sydney's Human Research Ethics Committee (protocol number 2018/130) approved this research study before the start of data collection.

Inclusion Criteria

To satisfy the inclusion criteria for participation in the study, participants were required to:

- be aged 12 years or above
- be either a young person attending a participating *headspace* center; a supportive other of a young person attending a participating *headspace* center (eg, family member, caregiver, friend); or a health professional, service manager, or administrator working at a participating *headspace* center
- be proficient in reading and speaking English
- complete the participant consent process.

Participants

headspace staff advertised this study using posters and postcards distributed throughout the participating centers. Recruitment was passive to avoid any perceived coercion, whereby a potential participant would contact the research officer listed on the advertisements to express their interest in the study and request further information. Upon the potential participant's request, the research officer would then forward (via email) on the participant information statement, participant consent form, and screening survey to determine eligibility.

For participants under the age of 15 years, both the young person and their guardian were given detailed and age-appropriate information about the research study before the workshop via a parental information statement and a child assent form. At the beginning of the workshop, research officers reminded all participants about what the workshop would involve, provided an opportunity to ask any questions, and reminded them that the participation was voluntary. The research officers spoke separately to participants under the age of 15 years and their guardians to ensure that they understood what the workshops would involve and what the study was about. They also

answered any questions and reminded the young person that they could withdraw from the study at any time without consequence. If a young person agreed to participate in the workshop, their guardian provided a signed parental consent form and the young person provided a signed child assent form.

The young people and their supportive others were reimbursed with a gift card valued at Aus $30 (US $23) for study participation.

Co-design Workshops

Workshop Location and Demographics

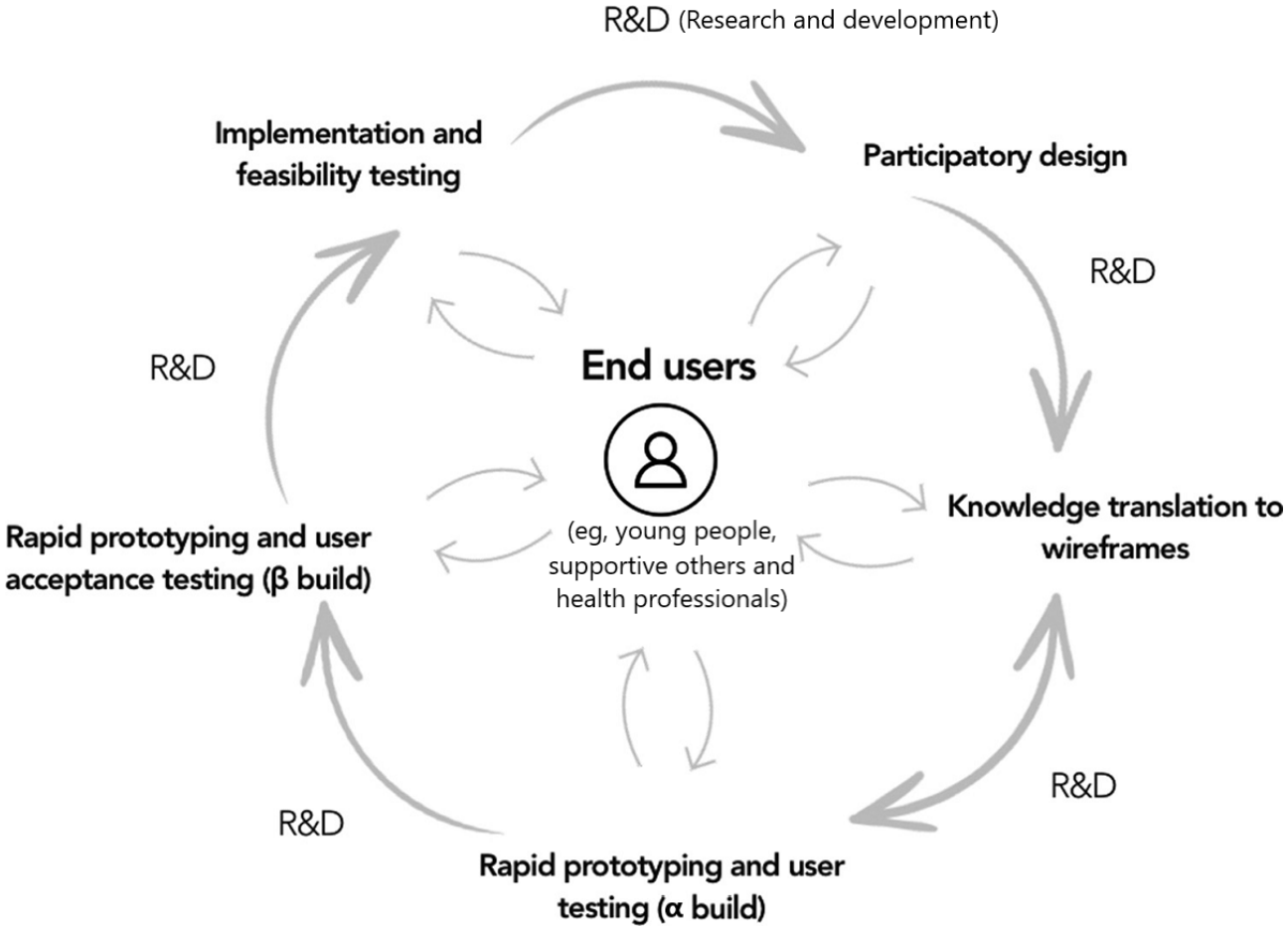
A total of 10 co-design workshops were held from July to September 2018 in *headspace* centers across the Australian states of New South Wales, South Australia, and Queensland.

We initially invited *headspace* centers to participate in this study, trying to ensure a representation of centers located in nonurban areas. Participating *headspace* centers were aware of the possibility of implementing this technology in their centers in the future. Although researchers did not specifically recruit Aboriginal and Torres Strait Islander people, LGBTQIA+ people, and CALD people, because of local population demographics, these groups were naturalistically represented in the participant sample.

Workshop Protocol

These workshops represent phase 1 of the Project Synergy Research and Development Cycle (Figure 1 [28]), where co-design workshops are rapidly conducted across different sites until theme saturation [28]. In these workshops, technology designs, ideas, and principles are evaluated by participants, sometimes iteratively (if enough time has passed for insights from a previous workshop to be translated into a testable wireframe or prototype). Previous work describing the process and outcomes of this methodology [28,29,34,35] as well as work on a separate mental health and well-being app [36] have been published.

Figure 1. Project synergy research and development cycle.



Although general workshop agendas were adhered to as appropriate, facilitators (the second, third, and fourth authors) also conducted workshops flexibly and followed up on topics in response to individual workshop dynamics. Certain topics were also given more priority to explore in different locations and contexts (eg, technology use and connectivity in nonurban areas). Therefore, the workshop content and outcomes varied slightly among workshops.

After each workshop, facilitators reviewed the workshop findings, adapting the general workshop agenda to remove data-saturated topics and add new and further topics of interest. An example of a workshop agenda is provided in Multimedia Appendix 1.

Owing to the number of participants and other contextual factors, workshops ranged from 2.5 to 4 hours in duration and consisted of the following stages: discovery, evaluation, and prototype. In the discovery stage, facilitators led discussion

around the following topics: general technology use, technology use for the purposes of supporting health and mental health, and internet use and connectivity. In the evaluation stage, participants were presented with paper printouts of various components of the BMC Youth Platform and asked to annotate them with their thoughts and comments. Finally, in the prototype stage, participants were asked to brainstorm new ideas, functionalities, and wireframes (with marker pens and sketchbooks) for the BMC Youth Platform. Owing to the sensitivity of the subject matter, workshops were not audiorecorded or videorecorded to decrease the risk of identification and facilitate participant disclosure. Instead, scribes took notes at each workshop. All workshops included at least one facilitator who was appropriately qualified in mental health to provide counseling support to the participants if needed.

Knowledge Translation

All workshop data, including notes and artifacts from the evaluation and prototype stages, were collated and reviewed by an independent knowledge translation team consisting of 2 young people (listed in the Acknowledgments) who had never previously been exposed to the BMC Youth Platform or its concepts. The knowledge translation team summarized the outcomes of the discovery and prototype stages and conducted a procedure similar to descriptive content analysis [37] on the outcomes of the evaluation stage. Specifically, all annotations were reviewed by each team member, who noted their general observations. They then coded the annotations together, organizing codes according to semantic themes representing different components of the BMC Youth Platform (Table 1).

Template Analysis

To supplement knowledge translation insights and identify general recommendations on designing HITs for mental health, we conducted a type of codebook thematic analysis [38] known as *template analysis* [39,40] on the workshop data. Specifically, the data set consisted of the knowledge-translated, summarized outcomes of the evaluation stage (in tally form) and scribe notes from each workshop. In recognition of the fact that generic thematic analysis (including template analysis) is a method with many different approaches that reflect a wide variety of epistemological (theory of knowledge) and ontological (theory of being) assumptions [38], we located our approach within a philosophical position of qualitative neopositivism. This position assumes a realist ontology and epistemology.

The data set was coded with NVivo 12 (QSR International). The first author initially coded all the data using a bottom-up, descriptive approach at a level close to the data (eg, *avoid clinical jargon* and *worry that information is confronting*). At the conclusion of this first round, the first author determined that certain groups of codes fit well with self-determination theory [41,42]. Although self-determination theory is originally a theory of intrinsic motivation, human-computer interaction research has found it to be applicable to user engagement with digital technologies [43,44]. Self-determination theory has also been successfully applied to user needs, facilitators, and barriers for mental health technologies [45,46].

Using self-determination theory as a partial reference for a priori themes, the first author grouped these descriptive codes into a preliminary coding template. A second coder (the second author) then independently coded 10% of the data according to this preliminary coding template to check its quality. In this case, quality was defined as the clarity of definitions and whether the template comprehensively covered the data set [40]. Both coders then discussed all discrepancies in coding until they were resolved, and insights from this process were used to refine the coding template. This iterative process of independent coding, comparison, and refinement of the coding template was repeated until a satisfactory level of interrater reliability was reached (Cohen kappa for all codes >0.65), which took 3 rounds of coding. Following this, the final interpretation of the template in relation to addressing our study aim was conducted.

Results

Participants

Table 2 reports workshop details, including dates, location, and participant characteristics. All workshops except workshop 5 (Ashfield, New South Wales) were conducted in nonurban areas. All *headspace* clients are young people aged between 16 and 25 years unless specified otherwise. Staff included mental health clinicians (psychologists, mental health nurses, and social workers), youth workers, administrative staff, and center managers from *headspace*.

Table 2. Workshop and participant details (N=105).

Workshop number	Workshop date	Location	Facilitators, n	Participants, n	Participant demographic information		
					Gender	Participant role	Cultural or personal identification[a]
1	June 18, 2018	Edinburgh North, South Australia	3	10	8 female and 2 male	6 clients, 1 support person, and 1 staff member	None disclosed
2	June 20, 2018	Edinburgh North, South Australia	3	11	7 female and 4 male	5 clients, 3 support people, and 3 staff members	1 CALD[b]
3	July 20, 2018	Broken Hill, New South Wales	3	9	4 female, 4 male, and 1 transgender	6 clients, 1 younger client aged between 12 and 15 years, and 2 staff members	2 CALD and 2 Aboriginal or Torres Strait Islander people
4	July 25, 2018	Townsville, Queensland	2	11	6 female and 5 male	8 clients and 3 staff members	2 Aboriginal or Torres Strait Islander people and 5 LGBTQIA+[c] people
5	August 9, 2018	Ashfield, New South Wales	3	5	4 female and 1 male	3 clients, 1 younger client aged between 12 and 15 years, and 1 support person	5 CALD
6	August 14, 2018	Wagga Wagga, New South Wales	2	9	7 female and 2 male	2 clients, 2 younger clients aged between 12 and 15 years, 3 staff members, and 2 support people	None disclosed
7	August 22, 2018	Bathurst, New South Wales	2	11	5 female and 6 male	6 clients, 4 younger clients aged between 12 and 15 years, and 1 staff member	None disclosed
8	August 28, 2018	Orange, New South Wales	2	22	6 female and 16 male	7 clients, 11 younger clients aged between 12 and 15 years, and 4 staff members	5 Aboriginal or Torres Strait Islander people
9	August 30, 2018	Wollongong, New South Wales	2	7	4 female, 2 male, and 1 gender neutral	6 clients and 1 staff member	1 person with disability
10	September 4, 2018	Dubbo, New South Wales	2	10	7 female, 2 male, and 1 gender neutral	7 clients and 3 staff members	4 LGBTQIA+

[a]Participant demographic information, particularly gender and cultural or personal identification, was provided by participants on a voluntary basis and was therefore unable to be captured consistently across workshops.

[b]CALD: culturally and linguistically diverse people.

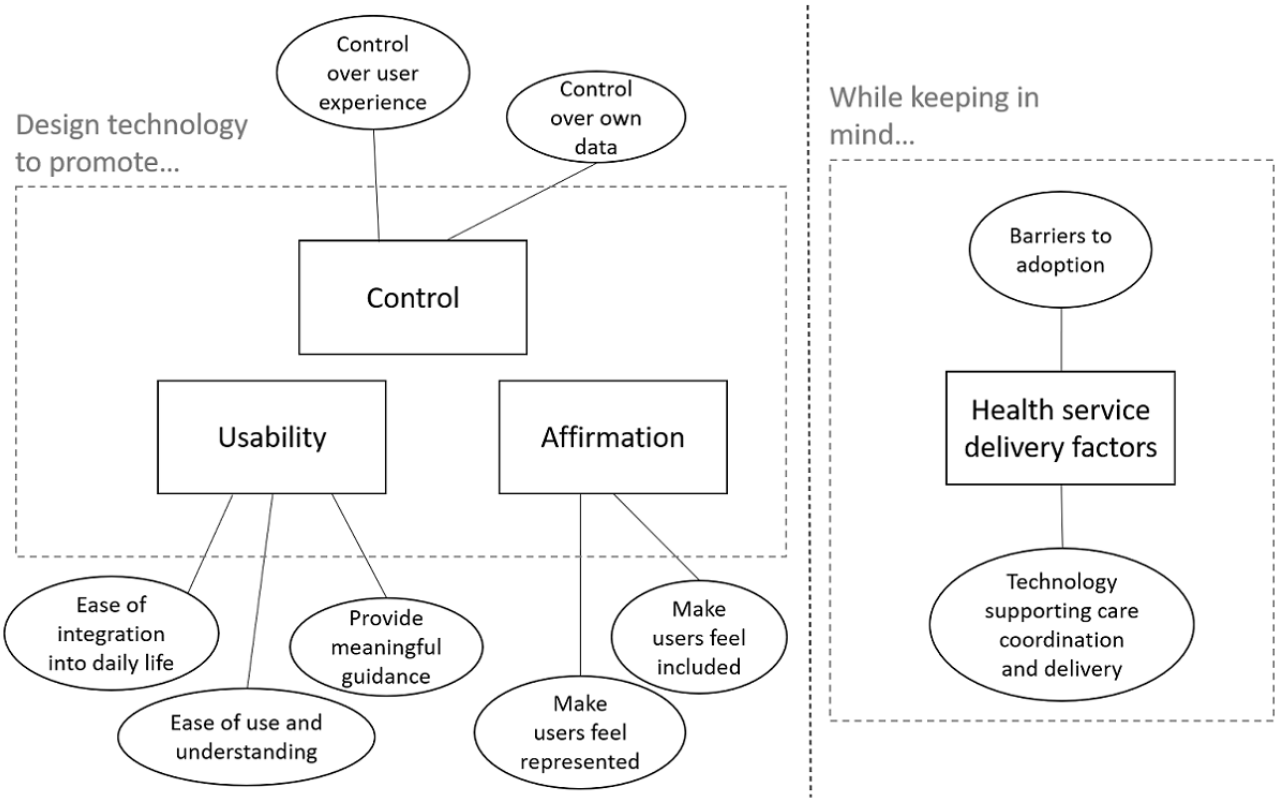
[c]People of diverse genders and sexualities, inclusive of and not limited to lesbian, gay, bisexual, transgender, queer, intersex, asexual, questioning, and pansexual people.

Template Analysis

Figure 2 shows the coding template following analysis and the refinement of the codes. During the analytical process, numerous codes were generated from the data set. These codes were organized into 11 low-level themes, which were then organized into the following high-level themes (design principles): control, usability, affirmation, and health service delivery factors. The full coding template (including a codebook of definitions) is presented in Multimedia Appendix 2.

Although many participants belonged to culturally diverse groups, they made many design recommendations applicable to universal human experience. These recommendations overlapped closely with the 3 basic psychological needs postulated by self-determination theory, namely, autonomy (corresponding to our theme of control), competence (corresponding to our theme of usability), and relatedness (corresponding to our theme of affirmation) [41,42].

Figure 2. Final coding template.



Give Users Control

First, participants emphasized the importance of having control over their user experience and their data.

Control Over User Experience

Participants made clear their expectations that any technology they use should give them the ability to control their own user experience, for example, by active customization of key design elements such as background color, layout, and the ability to upload an avatar. Participants from CALD backgrounds and participants with disabilities also expressed a desire to be able to change language and font size and to be able to consume information in different audiovisual formats (eg, text-to-speech options).

This desire for control also extended to the care being delivered to young people, with participants appreciating that they could make informed choices on which care options to embark on to progress their care. Participants also wanted to be able to arrange the health cards on their dashboard in ways that would make more sense or that they would find less overwhelming:

> *How the cards are ordered is important, [it] would be great to have the cards you want to work on together or first, group cards by color.* [Bathurst, young person]
>
> *Can you collapse aspects if [you're] feeling "shitty"?* [Wagga Wagga, young person]

Finally, participants also suggested providing mechanisms through which user feedback could be provided and actioned on, for example, through a dedicated feedback section.

Control Over Data

Participants also stressed the importance of having control over their own data at all times, not only through data security (authentication and encryption, eg, through password or personal identification numbers) but also through data privacy and sharing. A key feature of the BMC Youth Platform is that young persons' dashboards of results are visible to health professionals at their service. The young persons can also share their dashboards with trusted supportive others, such as parents or friends. Although participants found it acceptable to share their data in this form with their health professionals, they were more hesitant to share their data with their supportive others (eg, parents or friends). Participants wanted fine-grained control and the ability to grant or revoke their supportive others' access to certain health cards, particularly those deemed by young people to be more sensitive, such as substance and tobacco use:

> *I don't want mum to see smoking, but they can see anxiety.* [Edinburgh North, young person]

The consent and data sharing status of each of their health cards as well as other parts of their data (such as health history) would also need to be presented to the user clearly and transparently:

> *How long do they have access to the information? [It] needs expiry, for example when [you] repeat the questions, or between 1-30 days, can check in to add or remove people.* [Wollongong, young person]
>
> *Privacy setting on cards can be set with clinician at first appointment, then the parent or carer receives the email invitation. They might have their own dashboard to look at.* [Dubbo, headspace staff]

Make the Technology Usable

Second, participants expressed that the technology should not be difficult to use, that is, it should (1) be easy to use and understand, (2) provide meaningful guidance, and (3) be interoperable.

Ease of Use and Understanding

Participants specified that the technology should be easy to use and understand at first glance and that it should not be overwhelming or frustrating to use. The technology should be clear, unambiguous, and consistent in language and user interface (including layout, icons, and imagery), and clinical jargon should be avoided. The latter point was especially relevant for younger audiences aged between 12 and 15 years because of lower reading levels. However, although participants found clinical terms such as "social and occupational functioning" and "psychosis-like experiences" confusing and intimidating, it was difficult to come up with alternative, more acceptable terms. After considering the issue, many participants suggested using direct language and providing definitions.

Provide Meaningful Guidance

Participants further specified that the technology should present information in meaningful chunks or sequences (eg, displaying participants' data on the most important health domains first) to promote a gradual understanding of psychoeducational concepts. This was particularly the case for the onboarding stage, which was where many participants would encounter clinical terminology and their own data for the first time. In addition to presenting information in a staggered, chunked manner, the technology should also provide meaningful guidance on how to use and interpret the technology. Participants overwhelmingly endorsed a model of delivering information whereby a page was kept simple and clutter-free at first presentation (relating to the previous tenet of ease of use and understanding), with users being given the freedom to learn more via a *more information* button or by hovering over complex terms (eg, *mania*, *psychosis*, or *clinician*).

Similarly, participants expressed that the technology should provide prompts or mechanisms through which users could gain an understanding of how to use the technology and how the technology could help them, for example, via a help button or Frequently Asked Questions (FAQ) page or through prompts reminding them to check their dashboard and complete new psychometric questionnaires at appropriate time intervals. These prompts could be tailored based on their questionnaire responses:

> *If a health card is red, the health professional (or system) should keep close eye on this—and prompt you to repeat the questionnaire.* [Dubbo, headspace staff]

Interoperability

Participants also emphasized that any HIT should be easy to integrate into daily life. This technology should be able to compile data from other apps (eg, step tracking apps) and enable all information of a user to be accessed from one location. Similarly, some participants endorsed the idea of quick, convenient authentication via social media, although others were concerned about data security issues. Young people overwhelmingly expressed a preference for mobile usability (responsive web design) and mobile integration (eg, via SMS text messaging instead of email), as many of them had neither easy access to computers nor an email address.

Affirm Users' Identity and Preferences

Finally, participants emphasized the importance of an HIT making its users feel represented and valued.

Make Users Feel Represented

Despite one of the purposes of HITs being the facilitation of users' mental health care via routine outcome monitoring, participants (both health professionals and young people) felt that it should provide functionality to represent the user beyond traditional clinical information such as demographics and mental health status. Participants wanted to be able to define themselves in more nuanced ways and felt that the screenshots they were presented with reflected a technology that could not yet allow them to do so. Their suggestions for ways to define themselves comprehensively and accurately ranged from simple actions such as being able to specify their preferred pronoun and being offered a more diverse range of response options when specifying their demographics (eg, sexuality, gender, and ethnicity) to more comprehensive methods such as adding new health cards to provide more information on additional domains such as family violence and homelessness or elaborating on responses via free text. This was judged to be particularly relevant given that *headspace* also offers support for domains outside mental health (such as with work and study):

> *Dashboard needs to reflect what is important to me, [for example] culture, sexual health and homelessness.* [Dubbo, young person]
>
> *These questions are very mental health–focused but headspace does a lot more.* [Bathurst, headspace staff]

Some LGBTQIA+ participants also felt that their sexual identity and preferred pronouns should not "be hidden away" in a separate *About me* tab, as this made it harder for them to come out to their health professional, and that instead it should be reflected in the dashboard.

Participants also felt that HITs should cater to them. Not only should they be usable regardless of physical or mental ability (mental health status, disability, or other accessibility considerations) but they should also offer a personalized user experience and present their various components (questionnaires, care options, psychoeducation, language, etc) in ways that are tailored to their users' personal circumstances (eg, gender, culture, location, and questionnaire responses):

> *[It] should be designed like Spotify, when you select your music it also suggests other artist in the same sort of music. [On the BMC Youth Platform] if [you] choose tobacco it would also give you the other drug [questions] like alcohol and cannabis.* [Bathurst, young person]
>
> *Tabs at top [could link to psychoeducation resources]—possibly an option of "how to talk to my*

> *parents," "how to talk to my partner," [that is] some ideas about what you can do with this information in the meantime. It should be tailored to who the login "user" is, [that is] at a certain age it may be directed to something different than to a 13-year-old, or a parent.* [Edinburgh North]

Finally, the CALD participants emphasized that the technology should cater to their needs, namely, that their languages should be supported and that the technology should acknowledge that different cultures could approach mental health differently, for example, by viewing them as "life circumstances" instead of "mental illness."

Make Users Feel Valued

Both health professionals and young people were adamant that an HIT user should not feel reduced to a diagnosis or an assortment of diagnostic labels. Instead, the HIT should be designed to celebrate nonclinical aspects of personhood as well, such as a user's likes and aspirations (eg, via a *My Goals* section). Although these aspirations could relate to a user's mental health (eg, improving their mood or sleep), they could also reflect a user's other goals. Participants also suggested having a dedicated space for users to store personally meaningful resources, such as a journal, pictures of inspirational people, or self-care resources useful to them:

> *If clinicians can see this page [they] will have a more positive approach to the relationship [and] will be able to connect with the young person.* [Bathurst, young person]

Participants also raised concerns that the screenshots of the technology they were presented with could potentially frame users too negatively. Instead, they preferred a strengths-based, positive framing approach that provided positive affirmations, kept "users on track," and celebrated achievements and strengths. Participants also recommended adding gamification to the technology to increase motivation to engage and promote a sense of progress, for example, via markers of achievements (eg, ribbons or "becoming a BMC Youth Platform warrior") or a leveling and reward system. They further specified that any gamification should "reward effort more than outcome," that is, rewards should be assigned based on the level of long-term engagement and the extent to which the user has explored (or completed) the technology.

An HIT should also not alienate the user and make them feel alone in their struggles with mental health. Instead, the technology should promote a sense of social connection (even parasocial connection), for example, by providing testimonials and links to other people's mental health stories (eg, an "others that struggle with this issue" section) and by normalizing suboptimal mental health status (eg, "It's OK to feel bad").

Health Service Delivery Factors

In addition to the more universal design principles described above, participants also contributed insights into what could broadly be categorized as health service delivery factors. This included perceived barriers to the widespread adoption of HITs for mental health care and how HITs could be used to support mental health care coordination and delivery.

Barriers to Adopting HITs for Mental Health Care

Although this was not the key topic of investigation, during workshops, participants discussed barriers to adopting digital technologies for mental health care that were relevant to their regions and communities.

The limitations of internet connectivity in nonurban areas, raised in 7 of the 8 nonurban areas investigated, was the most commonly cited barrier, with slow internet speeds and a lack of access to home internet in certain geographic areas being the main issues. Participants were familiar with free Wi-Fi locations, such as libraries, local schools or universities, and fast-food restaurants. Using internet in nonurban areas was also associated with several costs that were unsustainable for young people, including the cost of purchasing smartphone apps as well as the cost of mobile data itself:

> *I use public Wi-Fi ([McDonald's] or university Wi-Fi) to get internet since my home connection is bad.* [Wagga Wagga, young person]

Health professionals who participated in the workshops also raised the issue of software fatigue. Currently, *headspace* protocols mandate multiple other pieces of software to be used to manage minimum data entry for its clients. Health professionals were hence cautious of new technologies, stating that "[it's a] real turn off to do the same questions more than once." Instead, they would prefer that client responses could be "shared between systems" in the case of overlap:

> *With 96 clients on my list [I] do not want 96 notifications via email as there is enough to review with the current system.* [Dubbo, headspace staff]

Technology Supporting Care Coordination and Delivery

Collectively, workshop participants outlined concrete suggestions through which HITs could support health professionals and clients during mental health care.

First, HITs could help health professionals and clients understand the progress of the client's care at all stages of the care. Before the first appointment, they could provide icebreakers to help potential clients become comfortable with their health professionals and provide health professionals with their client's dashboard of results to allow them to prepare for the first appointment more effectively. This was viewed as critically important given that at the time of the workshops, the Australian public mental health system only funded 6 1-hour sessions with a psychologist (extendible to 10 sessions given further referral) per year:

> *It would take the whole session to go over history alone, with only 10 sessions [we] can't afford to have a session devoted to not achieving anything.* [Wollongong, young person #1]

> *We need something [that isn't] telling your story again to a new psychologist each time. It's not schema therapy with 20 sessions, 6 sessions are precious.* [Wollongong, young person #2]

During the course of care, HITs could maintain a record of the client's health over time and over a variety of domains (multidimensional routine outcome monitoring) as well as a

record of care options attempted over time. This record could then be used as a reference to prompt topics of discussion between the health professional and client:

> *I want to see my apps linked into the system to share with headspace, so they can see my progress in real time.* [Dubbo, young person]
>
> *[On the system] buttons could say "what can I do now?," "what to chat to my clinician about?"* [Wagga Wagga]

This record could be exported and transferred over to a new mental health service or a new health professional, should the client require it, to minimize administrative and psychological burden on the young person interacting with their new health professional:

> *Retelling your story is one of the hardest [things].* [Wollongong, young person]

By enabling tracking of a client's progress during care, HITs could also support the health professional and client in gaining a more unified understanding of how to direct the client's care. The health professional could tailor their support according to their clients' needs in mental health and other domains. Similarly, by seeing how their self-reported health changed over time and following different care options, the client could make more informed choices on which health domains and care options to focus on.

Such technologies could also improve the efficiency and safety of clinical care. For example, video calling functionalities could overcome long distances in large rural health catchments and allow for more flexible appointment scheduling. Young people also expressed that they would find it easier to complete questionnaires on a device rather than answer them face-to-face with a health professional. Finally, such technologies could alert mental health services when their clients are experiencing emergency scenarios in real time, assisting them in risk management.

Discussion

Using Self-Determination Theory to Design HITs for Mental Health

The findings of this study can be broadly categorized into (1) design recommendations for HITs for mental health and well-being and (2) health service delivery factors to consider when designing such technologies. In other words, the former category reflects how such technologies should be *designed* and the latter category reflects how such technologies should be *used*.

The design recommendations approximately map to the 3 basic psychological needs proposed by self-determination theory: autonomy, competence, and relatedness [41]. As mentioned before, self-determination theory is a theory of motivation (including both intrinsic and extrinsic motivation) that has been applied to the design of health technologies [44,47] as well as web-based help seeking for mental health problems [45,46]. The fact that we found our workshop findings to align with self-determination theory constructs is a further support for its application to the design and delivery of HITs. Currently, many HITs do not explicitly apply theories of engagement or health behavior change to their design and delivery [48]. However, our results suggest that the contribution of these theories to HITs is equally important as that of evidence-based clinical content.

Importantly, although our core design recommendations map neatly onto the 3 basic psychological needs, how they are executed in practice can simultaneously support multiple needs [46]. For example, from our data, we found technological accessibility to be a blend of *usability* (ease of use), *control* (eg, customizing visual elements such as font size), and *affirmation* (inclusiveness) and avoiding clinical jargon to be a blend of *usability* (not using difficult terminology) and *affirmation* (avoiding making users feel reduced to a diagnosis). A comprehensive list of recommended design features identified from our workshops is shown in Table 3.

Our results are broadly consistent with previous research, reflecting the importance of promoting social connection [32,46], personalization and customization [36,46], clear and casual language [36], and addressing data security and privacy concerns [7,49].

Our results also suggest that any recommendations from participants should be evaluated through the lens of self-determination theory, with explicit consideration given to how each suggested change could promote each of the basic psychological needs. This is in line with previous research on applying self-determination theory to help seeking for mental health problems [45,46]. This measured evaluation is important given that many participant suggestions involved broad concepts or specifications and may reveal that further expertise needs to be consulted. For example, participants recommended implementing gamification to "reward effort more than outcome" and suggested several mechanics through which HITs could be gamified, including level-based reward systems and achievements. However, gamification is not merely a collection of gamification elements but instead the deliberate integration of gameful mechanics into a technology to support its core functionalities [50]. Implementing this suggestion, therefore, would involve not only the iterative user testing of the gamified HIT with the target end user population but also the consultation of gamification designers to integrate gamification at a deeper, systemic level that promotes both the HIT's aims and user motivation to engage with the HIT (via the 3 basic psychological needs).

Table 3. Participants' recommended design features and corresponding basic psychological needs.

Broad design category	Recommended design feature	Supported basic psychological needs
User interface and experience	Customizable user interface (language, colors, layout, font size, etc) to accommodate user preferences and needs	Autonomy and competence
User interface and experience	Tailored user experience	Relatedness
User interface and experience	Meets international accessibility standards (eg, web content accessibility guidelines)	Competence and relatedness
User interface and experience	Customizable information input and output (eg, being able to submit free text that more accurately describes you and being able to organize how your data are displayed) for ease of tracking and understanding	Autonomy, competence, and relatedness
User interface and experience	Clear layout, icons, and imagery	Competence
Content and functionality	Ability to provide user feedback that will be actioned on (eg, feedback or evaluation section)	Autonomy and relatedness
Content and functionality	Instructional prompts giving guidance on what to expect and how the technology will help the user and reminder prompts to promote reengagement	Competence
Content and functionality	Provision of optional additional information not core to the experience; however, the user can consult to learn more (eg, psychoeducation or a frequently asked questions section)	Autonomy and competence
Content and functionality	Gradual onboarding and the provision of information in meaningful chunks to support learning and understanding (eg, showing the most important health domains first on a health dashboard)	Competence
Content and functionality	Consider element of fun (eg, gamification)	Relatedness
Content and functionality	Promote social connection with peers and communicate to users that they are not alone (eg, through peer support groups and testimonials)	Relatedness
Language and tone	Adopt a strengths-based approach and celebrate nonclinical aspects of personhood such as likes, aspirations, strengths, and achievements	Relatedness
Language and tone	Cultural competence	Relatedness
Language and tone	Clear, casual, unambiguous, and consistent language that avoids clinical jargon and loaded terms	Competence and relatedness
Interoperability	Ease of integration with other apps and technologies (eg, health apps or convenient authentication methods)	Competence
Interoperability	Mobile integration as young people do not use email or computers frequently	Autonomy and competence
Security and privacy	Industrial-grade data security	Autonomy
Security and privacy	Fine-grained (individual level) data sharing functionality	Autonomy

Often, participant recommendations seemed to contradict each other at the surface level. For example, we observed tensions within participant recommendations on language, with an emphasis on increasing understanding (eg, by using clinically accepted terms and providing definitions when needed) conflicting with an emphasis on using casual, simple, and nonloaded language. Similarly, the focus on ease of use (by presenting information clearly and in a manner that is not overwhelming) superficially conflicted with the desire to have access to contextualizing information as well as the desire to improve representation by increasing the number of psychosocial domains represented on the dashboard of results. When this contradiction was raised, participants proposed the solution of giving users the ability to customize their own user experience and make it as simple, or as complex, as they liked. Under self-determination theory, this solution would allow each user to satisfy their needs for autonomy, competence, and relatedness at an individual level.

However, although many contradictions were solved in a way that was acceptable to our participants on an individual level, some solutions raise implications on the service level. For example, an overarching recommendation from the workshops that participants overwhelmingly agreed on was the desired ability to provide more nuanced information about oneself via free text. On the service level, however, this solution would raise implications for, and potentially conflict against, clinical safety protocols unless the free-text fields were moderated and constantly monitored (on a 24-hour basis) for any indication of suicide risk. Currently, the role of this moderator would most practically be filled by a health professional employed at the service, a human resource allocation luxury many publicly funded mental health services do not have. Although it is important to adhere to clinical safety protocols, this situation is an example of a dilemma faced by decision makers where design practices that can best serve the intended user base are disincentivized by the wider public health and medical research systems, resulting, in some respects, in a lower-quality HIT.

The complexities and often contradicting incentives in the public health and medical research spheres can have the effect of frustrating progress in what should be the shared aim of improving public health outcomes. In such cases, marginalized populations that face multiple compounding barriers to access to quality care are the ones that are the most adversely affected, particularly in acute, abnormal situations such as natural disasters or a global pandemic [51]. To achieve the most impact, HIT solutions and software development resources should directly solve problems such as this that disproportionately impact marginalized populations. For example, one method of overcoming this problem could be to automate content moderation via natural language processing [52].

Designing for Culturally Diverse Groups

The findings of this study suggest that although culturally diverse groups have specific concerns about design (eg, affirmation of diversity in sexual preference and gender identity and the provision of multiple language options), their needs fall broadly within universal design principles. For example, our LGBTQIA+ participants wanted a way to display their sexuality and pronouns alongside their dashboard of results. In addition to being able to represent themselves more fully, this would also increase their comfort levels with their health professional as they would not have to directly disclose this information face-to-face. The psychosocial context behind our participants' recommendations is crucial and cannot be divorced from the recommendations themselves without running the risk of developing solutions that could alienate the culturally diverse groups from which they hail. For this reason, we have synthesized our findings into general rather than specific recommendations.

Textbox 1 provides general recommendations on designing HITs for mental health that were drawn from co-design with culturally diverse populations and based on self-determination theory. Specifically, designers should consider the intended functionality of the HIT and how the HIT can fulfill its users' basic needs of autonomy, competence, and relatedness (by providing perceptions of control, usability, and affirmation). This would require identifying barriers to their access to, and adoption of, the technology (ideally through a co-design process similar to what was adopted in this study) and designing the technology to mitigate these barriers.

Textbox 1. General recommendations on designing health information technologies for mental health based on self-determination theory.

1. Identify the purpose of the health information technology (HIT) and the mechanisms through which it achieves it aims
2. Consider how the HIT can fulfill users' basic needs of autonomy, competence, and relatedness through providing perceptions of control, usability, and affirmation:
   - Control includes the ability to control user experience as well as the ability to control user data
   - Usability includes ease of use and understanding, the provision of meaningful guidance, and ease of integration into daily life (eg, interoperability)
   - Affirmation includes users feeling both valued and included within the target user base
3. Consider the possible barriers to adopting the HIT, including access and use, from all potential user groups
4. Consider how to design the HIT to directly address as many adoption barriers as possible while still preserving perceptions of control, usability, and affirmation

Adoption barriers that this study and previous research have identified include young people's lack of easy access to personal computers [53] as well as poor internet connectivity in many nonurban parts of Australia [54]. HIT designers should be aware of inequalities and how they compound and not perpetuate them. For example, they should neither enable discrimination nor systematically exclude groups of users who may not have adequate internet connectivity or reading level. Compassionate design that puts these barriers at the forefront of consideration is required to ensure that technologies can be accessed by all. Our study results underscore the importance of user testing with the target population to confirm that the technology satisfies their needs and to identify potential barriers that may prevent them from accessing these technologies.

Limitations

A limitation of this study was that the researchers did not exclusively advertise to and recruit participants from the populations of interest within this study (Aboriginal and Torres Strait Islander people, LGBTQIA+ people, and CALD people in nonurban areas). Instead, all young people attending *headspace* (aged between 12 and 25 years), *headspace* staff, and supportive others were invited to participate in the workshops. Therefore, although participants who identified as being of these populations were present at workshops, the data collected cannot be exclusively attributed to these populations. Future research should consider recruiting participants exclusively from these populations and provide appropriate support to ensure comfort in communication (eg, an Aboriginal or Torres Strait Islander person to facilitate the workshop tailored to that population). Artifacts from the *headspace* Dubbo workshop (annotated screenshots and drawings of technology prototypes) were not included in the analysis for this study, but all researcher-scribed notes were included.

Our inclusion criteria would also have excluded CALD participants who were not proficient in English. Hence, our study results cannot be fully generalized to CALD communities. As mentioned above, future research should recruit exclusively from CALD populations and provide appropriate support (eg, by conducting workshops in participants' native language or providing a translator).

Furthermore, our workshops included a mixed group of participants (young people, their supportive others, and health professionals). This gave facilitators the opportunity to address tensions and potential contradictions in what each of these groups wanted from an HIT for mental health. However, although workshop facilitators did their best to encourage open discussion, the implicit power difference among these groups could have discouraged this from participants belonging to less socially powerful groups (eg, young people). Similarly, workshops were not audiorecorded or videorecorded to facilitate discussion. However, this limited the details on which scribe notes could be taken.

Finally, these insights arise from a relatively narrow (though important) source, namely, representative end users, including young people, health professionals, and supportive others (our participants), further interpreted through the lens of independent knowledge translators (2 young Australian women) and mental health researchers (the authors). Although these groups have their own spheres of expertise, they are not the only important perspectives in the design of HITs for mental health. There are other considerations to account for in the design of such technologies, particularly in terms of specialist detail in user experience design, security, governance, and software engineering. As a result, the findings of this study should be taken as informed recommendations for how similar HITs can be structured as well as a validation of existing ideas only.

Conclusions

Recent research has underscored the importance of iteratively testing the acceptability of HITs with the target end user population [25,33]. Through a series of co-design workshops, this study sought to extend the scope of our previous research from urban and nonurban Australian adolescents to include culturally diverse populations in nonurban areas of Australia. Although we identified several barriers and preferences specific to these populations in our co-design workshops, our results support the application of theory-based design of HITs (eg, self-determination theory) to develop user experiences that fulfill the universal basic psychological needs of autonomy, competence, and relatedness, through providing perceptions of control, usability, and affirmation. Deeper reflection on our findings also reveals the inherent tensions and difficulties in balancing the multiple, sometimes contradicting requirements of mental health technology stakeholders, including, but not limited to, health professionals, regulatory bodies, individual users, service managers, and best practice of the software development industry.

Acknowledgments

The authors would like to acknowledge Future Generation Global for funding the Youth Mental Health and Technology Program, which aims to improve young people's access to quality mental health care. The authors also thank Jessica Cottle for her assistance in preparing the data for analysis and Anna Roberts and Hannah Yee for knowledge translation.

Conflicts of Interest

IH was an inaugural commissioner on Australia's National Mental Health Commission (2012-2018). He is the Codirector, Health and Policy, at the University of Sydney's BMC. The BMC operates an early intervention youth service at Camperdown under contract to headspace. He is the chief scientific advisor to and a 5% equity shareholder in InnoWell Pty Ltd. InnoWell was formed by the University of Sydney (45% equity) and PricewaterhouseCoopers (Australia; 45% equity) to deliver the Aus $30 (US $23.23) million Australian Government–funded Project Synergy (2017-2020; a 3-year program for the transformation of mental health services) and to lead the transformation of mental health services internationally through the use of innovative technologies. At time of publication, VC is now employed at the Australian Digital Health Agency. TD is now the Director (Research and Evaluation), Design and Strategy division, at the Australian Digital Health Agency. None of the other authors declare any conflicts of interest.

Multimedia Appendix 1
Example workshop agenda.
[DOCX File , 80 KB - jmir_v23i2e23502_app1.docx]

Multimedia Appendix 2
Full coding template and codebook of definitions.
[DOCX File , 22 KB - jmir_v23i2e23502_app2.docx]

References

1. Bishop L, Random A, Laverty M, Gale L. Mental health in remote and rural communities. Royal Flying Doctor Service of Australia. Canberra: Royal Flying Doctor Service of Australia; 2017. URL: https://www.flyingdoctor.org.au/assets/documents/RN031_Mental_Health_D5.pdf [accessed 2020-08-15]
2. Lawrence D, Johnson SE, Hafekost J, Boterhoven de Haan K, Sawyer MG, Ainley J, et al. The Mental Health of Children and Adolescents. Report on the second Australian Child and Adolescent Survey of Mental Health and Wellbeing. Department

of Health. Canberra: Department of Health; 2016 Sep. URL: https://www1.health.gov.au/internet/main/publishing.nsf/content/9DA8CA21306FE6EDCA257E2700016945/$File/child2.pdf [accessed 2020-07-21]
3. Brenes GA, Danhauer SC, Lyles MF, Hogan PE, Miller ME. Barriers to Mental Health Treatment in Rural Older Adults. Am J Geriatr Psychiatry 2015 Nov;23(11):1172-1178 [FREE Full text] [doi: 10.1016/j.jagp.2015.06.002] [Medline: 26245880]
4. Fennell K, Hull M, Jones M, Dollman J. A comparison of barriers to accessing services for mental and physical health conditions in a sample of rural Australian adults. Rural Remote Health 2018 Feb;18(1):4155 [FREE Full text] [doi: 10.22605/RRH4155] [Medline: 29451985]
5. Seidler ZE, Rice SM, Dhillon HM, Cotton SM, Telford NR, McEachran J, et al. Patterns of Youth Mental Health Service Use and Discontinuation: Population Data From Australia's Headspace Model of Care. Psychiatr Serv 2020 Nov 01;71(11):1104-1113. [doi: 10.1176/appi.ps.201900491] [Medline: 32790590]
6. NMHC. The National Review of Mental Health Programmes and Services. National Mental Health Commission. Sydney: National Mental Health Commission; 2014. URL: https://www.mentalhealthcommission.gov.au/Monitoring-and-Reporting/national-reports/2014-Contributing-Lives-Review [accessed 2020-08-02]
7. Brown A, Rice SM, Rickwood DJ, Parker AG. Systematic review of barriers and facilitators to accessing and engaging with mental health care among at-risk young people. Asia-Pacific Psychiatry 2015 Aug 03;8(1):3-22. [doi: 10.1111/appy.12199]
8. Hilferty F, Cassells R, Muir K, Duncan A, Christensen D, Mitrou F. Is headspace making a difference to young people's lives? Final Report of the independent evaluation of the headspace program. Social Policy Research Centre, UNSW Australia. Sydney: Social Policy Research Centre, UNSW Australia; 2015. URL: https://headspace.org.au/assets/Uploads/Evaluation-of-headspace-program.pdf [accessed 2020-08-02]
9. Rosenstreich G. LGBTI People: Mental Health and Suicide. Revised 2nd Edition. National LGBTI Health Alliance. Sydney: National LGBTI Health Alliance; 2013. URL: https://www.beyondblue.org.au/docs/default-source/default-document-library/bw0258-lgbti-mental-health-and-suicide-2013-2nd-edition.pdf?sfvrsn=2 [accessed 2020-08-02]
10. Hunter E. Disadvantage and discontent: A review of issues relevant to the mental health of rural and remote Indigenous Australians. Aust J Rural Health 2007 Apr;15(2):88-93. [doi: 10.1111/j.1440-1584.2007.00869.x] [Medline: 17441816]
11. Zubrick S, Silburn S, Lawrence D, Mitrou F, Dalby R, Blair E. The Western Australian Aboriginal Child Health Survey: The Social and Emotional Wellbeing of Aboriginal Children and Young People. In: The Western Australian Aboriginal Child Health Survey. Perth: Curtin University of Technology and Telethon Institute for Child Health Research; 2005.
12. Wohler Y, Dantas JA. Barriers Accessing Mental Health Services Among Culturally and Linguistically Diverse (CALD) Immigrant Women in Australia: Policy Implications. J Immigr Minor Health 2017 Jun;19(3):697-701. [doi: 10.1007/s10903-016-0402-6] [Medline: 27002625]
13. Cross W, Singh C. Dual vulnerabilities: Mental illness in a culturally and linguistically diverse society. Contemporary Nurse 2014 Dec 17;42(2):156-166. [doi: 10.5172/conu.2012.42.2.156]
14. Kuipers P, Lindeman MA, Grant L, Dingwall K. Front-line worker perspectives on Indigenous youth suicide in Central Australia: initial treatment and response. Advances in Mental Health 2016 Mar 15;14(2):106-117. [doi: 10.1080/18387357.2016.1160753]
15. Hinton R, Kavanagh DJ, Barclay L, Chenhall R, Nagel T. Developing a best practice pathway to support improvements in Indigenous Australians' mental health and well-being: a qualitative study: Figure 1. BMJ Open 2015 Aug 27;5(8):e007938. [doi: 10.1136/bmjopen-2015-007938]
16. Kilian A, Williamson A. What is known about pathways to mental health care for Australian Aboriginal young people?: a narrative review. Int J Equity Health 2018 Jan 27;17(1):12 [FREE Full text] [doi: 10.1186/s12939-018-0727-y] [Medline: 29374482]
17. Smith E, Jones T, Ward R, Dixon J, Mitchell A, Hillier L. From Blues to Rainbows: Mental health and wellbeing of gender diverse and transgender young people in Australia. In: he Australian Research Centre in Sex, Health, and Society. Melbourne: The Australian Research Centre in Sex, Health, and Society; 2014:9781-9219.
18. Robinson K, Bansel P, Denson N, Ovenden G, Davies C. Growing Up Queer: Issues Facing Young Australians Who Are Gender Variant and Sexuality Diverse. Young and Well Cooperative Research Centre. Melbourne: Young and Well Cooperative Research Centre; 2014. URL: https://www.twenty10.org.au/wp-content/uploads/2016/04/Robinson-et-al.-2014-Growing-up-Queer.pdf [accessed 2020-08-02]
19. Minas H, Kakuma R, Too L, Vayani H, Orapeleng S, Prasad-Ildes R, et al. Mental health research and evaluation in multicultural Australia: developing a culture of inclusion. Int J Ment Health Syst 2013;7(1):23. [doi: 10.1186/1752-4458-7-23]
20. de Anstiss H, Ziaian T, Procter N, Warland J, Baghurst P. Help-seeking for Mental Health Problems in Young Refugees: A Review of the Literature with Implications for Policy, Practice, and Research. Transcult Psychiatry 2009 Dec 21;46(4):584-607. [doi: 10.1177/1363461509351363]
21. Forbes-Mewett H, Sawyer A. Mental health issues amongst international students in Australia: Perspectives from professionals at the coal-face. In: Threadgold S, Kirby E. editors. Australia: The Australian Sociological Association (TASA); 2011:1-19.
22. Thompson T, Brailer D. The Decade of Health Information Technology: Delivering Consumer-centric and Information-rich Health Care. Framework for Strategic Action. Washington, D.C: Department of Health & Human Services; 2004. URL:

http://www.providersedge.com/ehdocs/ehr_articles/the_decade_of_hit-delivering_customer-centric_and_info-rich_hc.pdf [accessed 2020-12-23]
23. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014 Aug;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
24. Hickie IB, Davenport TA, Burns JM, Milton AC, Ospina-Pinillos L, Whittle L, et al. Project Synergy: co-designing technology-enabled solutions for Australian mental health services reform. Med J Aust 2019 Oct;211 Suppl 7:S3-S39. [doi: 10.5694/mja2.50349] [Medline: 31587276]
25. Fleming TM, de BD, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the Impact of e-Therapy and Serious Gaming: Time for a Paradigm Shift. Front Psychiatry 2016;7:65 [FREE Full text] [Medline: 27148094]
26. Pretorius C, Chambers D, Cowan B, Coyle D. Young People Seeking Help Online for Mental Health: Cross-Sectional Survey Study. JMIR Ment Health 2019 Aug 26;6(8):e13524. [doi: 10.2196/13524]
27. Rowe SC, Davenport TA, Easton MA, Jackson TA, Melsness J, Ottavio A, et al. Co‐designing the InnoWell Platform to deliver the right mental health care first time to regional youth. Aust. J. Rural Health 2020 Apr 12;28(2):190-194. [doi: 10.1111/ajr.12617]
28. Hickie IB, Davenport TA, Burns JM, Milton AC, Ospina-Pinillos L, Whittle L, et al. Project Synergy: co-designing technology-enabled solutions for Australian mental health services reform. Med J Aust 2019 Oct;211 Suppl 7:S3-S39. [doi: 10.5694/mja2.50349] [Medline: 31587276]
29. Ospina-Pinillos L, Davenport TA, Ricci CS, Milton AC, Scott EM, Hickie IB. Developing a Mental Health eClinic to Improve Access to and Quality of Mental Health Care for Young People: Using Participatory Design as Research Methodologies. J Med Internet Res 2018 May 28;20(5):e188 [FREE Full text] [doi: 10.2196/jmir.9716] [Medline: 29807878]
30. Hagen P, Collin P, Metcalf A, Nicholas M, Rahilly K, Swainston N. Participatory Design of evidence-based online youth mental health promotion, prevention, early intervention and treatment. Melbourne: Young and Well Cooperative Research Centre; 2012:978.
31. Yarosh S, Schueller SM. "Happiness Inventors": Informing Positive Computing Technologies Through Participatory Design With Children. J Med Internet Res 2017 Jan 17;19(1):e14 [FREE Full text] [doi: 10.2196/jmir.6822] [Medline: 28096066]
32. Jessen S, Mirkovic J, Ruland CM. Creating Gameful Design in mHealth: A Participatory Co-Design Approach. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11579 [FREE Full text] [doi: 10.2196/11579] [Medline: 30552080]
33. Sockolow P, Schug S, Zhu J, Smith T, Senathirajah Y, Bloom S. At-risk adolescents as experts in a new requirements elicitation procedure for the development of a smart phone psychoeducational trauma-informed care application. Informatics for Health and Social Care 2016 Jun 03;42(1):77-96. [doi: 10.1080/17538157.2016.1177532]
34. Ospina-Pinillos L, Davenport T, Mendoza Diaz A, Navarro-Mancilla A, Scott EM, Hickie IB. Using Participatory Design Methodologies to Co-Design and Culturally Adapt the Spanish Version of the Mental Health eClinic: Qualitative Study. J Med Internet Res 2019 Aug 02;21(8):e14127. [doi: 10.2196/14127]
35. Ospina-Pinillos L, Davenport T, Navarro-Mancilla A, Cheng V, Cardozo AA, Rangel A. Involving end users in adapting the Spanish version of the Mental Health eClinic for young people in Colombia: A pilot study using participatory design methodologies. JMIR Preprint posted online September 01, 2019. [doi: 10.2196/preprints.15914]
36. Cheng VWS, Davenport TA, Johnson D, Vella K, Mitchell J, Hickie IB. An App That Incorporates Gamification, Mini-Games, and Social Connection to Improve Men's Mental Health and Well-Being (MindMax): Participatory Design Process. JMIR Ment Health 2018 Nov 19;5(4):e11068. [doi: 10.2196/11068]
37. Bauer M, Gaskell G. Qualitative Researching with Text, Image and Sound. London: SAGE Publications Ltd. London: SAGE Publications Ltd; 2000. URL: http://methods.sagepub.com/book/qualitative-researching-with-text-image-and-sound [accessed 2020-12-23]
38. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
39. King N, Brooks J, Tabari S. Template Analysis in Business Management Research. In: Qualitative Methodologies in Organization Studies: Volume II: Methods and Possibilities. Cham: Springer International Publishing; 2018:179-206.
40. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
41. Deci EL, Ryan RM. The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. Psychological Inquiry 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
42. Ryan R, Deci E. Overview of self-determination theory: An organismic dialectical perspective. Handbook of self-determination research 2002:A. [doi: 10.4135/9781446249215.n21]
43. Przybylski AK, Rigby CS, Ryan RM. A motivational model of video game engagement. Review of General Psychology 2010;14(2):154-166. [doi: 10.1037/a0019440]
44. Ryan RM, Rigby CS, Przybylski A. The Motivational Pull of Video Games: A Self-Determination Theory Approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]

45. Pretorius C, Chambers D, Coyle D. Young People's Online Help-Seeking and Mental Health Difficulties: Systematic Narrative Review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
46. Pretorius C, McCashin D, Kavanagh N, Coyle D. Searching for Mental Health: A Mixed-Methods Study of Young People's Online Help-seeking. In: CHI '20. New York: Association for Computing Machinery; 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April, 2020; Honolulu HI USA p. A. [doi: 10.1145/3313831.3376328]
47. Cheek C, Fleming T, Lucassen MF, Bridgman H, Stasiak K, Shepherd M, et al. Integrating Health Behavior Theory and Design Elements in Serious Games. JMIR Mental Health 2015 Apr 21;2(2):e11. [doi: 10.2196/mental.4133]
48. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
49. Mayer G, Gronewold N, Alvarez S, Bruns B, Hilbel T, Schultz J. Acceptance and Expectations of Medical Experts, Students, and Patients Toward Electronic Mental Health Apps: Cross-Sectional Quantitative and Qualitative Survey Study. JMIR Ment Health 2019 Nov 25;6(11):e14018. [doi: 10.2196/14018]
50. Deterding S. The Lens of Intrinsic Skill Atoms: A Method for Gameful Design. Human–Computer Interaction 2015 May 15;30(3-4):294-335. [doi: 10.1080/07370024.2014.993471] [Medline: 26281194]
51. Bowleg L. We're Not All in This Together: On COVID-19, Intersectionality, and Structural Inequality. Am J Public Health 2020 Jul;110(7):917-917. [doi: 10.2105/ajph.2020.305766]
52. Hussain MS, Li J, Ellis LA, Ospina-Pinillos L, Davenport TA, Calvo RA, et al. Moderator Assistant: A Natural Language Generation-Based Intervention to Support Mental Health via Social Media. Journal of Technology in Human Services 2015 Dec 14;33(4):304-329. [doi: 10.1080/15228835.2015.1105768]
53. Ellis LA, Collin P, Davenport TA, Hurley PJ, Burns JM, Hickie IB. Young men, mental health, and technology: implications for service design and delivery in the digital age. J Med Internet Res 2012 Nov 22;14(6):e160 [FREE Full text] [doi: 10.2196/jmir.2291] [Medline: 23171827]
54. Orlowski S, Lawn S, Antezana G, Venning A, Winsall M, Bidargaddi N, et al. A Rural Youth Consumer Perspective of Technology to Enhance Face-to-Face Mental Health Services. J Child Fam Stud 2016 Jun 18;25(10):3066-3075. [doi: 10.1007/s10826-016-0472-z]

Abbreviations

BMC: Brain and Mind Centre
CALD: culturally and linguistically diverse
HIT: Health Information Technology
LGBTQIA+ people: People of diverse genders and sexualities, inclusive of and not limited to lesbian, gay, bisexual, transgender, queer, intersex, asexual, questioning, and pansexual people
YMH and Tech Program: Youth Mental Health and Technology Program

Edited by G Eysenbach, R Kukafka; submitted 14.08.20; peer-reviewed by C Pretorius, A Čuš; comments to author 29.09.20; revised version received 14.10.20; accepted 12.11.20; published 10.02.21.

Please cite as:
Cheng VWS, Piper SE, Ottavio A, Davenport TA, Hickie IB
Recommendations for Designing Health Information Technologies for Mental Health Drawn From Self-Determination Theory and Co-design With Culturally Diverse Populations: Template Analysis
J Med Internet Res 2021;23(2):e23502
URL: https://www.jmir.org/2021/2/e23502
doi:10.2196/23502
PMID:33565985

Original Paper

Durability of Abstinence After Completing a Comprehensive Digital Smoking Cessation Program Incorporating a Mobile App, Breath Sensor, and Coaching: Cohort Study

Jennifer D Marler[1], MD; Craig A Fujii[1], BS, MIDS; Joseph A Galanko[2], PhD; Daniel J Balbierz[1], BS; David S Utley[1], MD

[1]Carrot Inc., Redwood City, CA, United States

[2]Biostatistics Core for the Center for Gastrointestinal Biology and Disease and the Clinical Nutrition Research Center, Department of Medicine, Division of Gastroenterology and Hepatology, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Jennifer D Marler, MD
Carrot Inc.
1400A Seaport Blvd
Suite 501
Redwood City, CA, 94063
United States
Phone: 1 415 757 7696
Email: marler@carrot.co

Abstract

Background: Despite decreasing prevalence over the last several decades, cigarette smoking remains the leading cause of preventable death and disease, underscoring the need for innovative, effective solutions. Pivot is a novel, inclusive smoking cessation program designed for smokers along the entire spectrum of readiness to quit. Pivot leverages proven methods and technological advancements, including a personal portable breath carbon monoxide sensor, smartphone app, and in-app text-based coaching. We previously reported outcomes from the end of active Pivot program participation in 319 adult smokers. Herein, we report longer-term follow up in this cohort.

Objective: The aim of this study was to assess and report participant outcomes 3 months after completion of Pivot, including smoking behavior, quit rates, continuous abstinence rates and durability, and predictors of abstinence.

Methods: This prospective remote cohort study included US-based cigarette smokers aged 18 to 65 years who smoked ≥5 cigarettes per day (CPD). Three months after completion of active participation in Pivot, final follow-up data were collected via an online questionnaire. Outcomes included smoking behavior (CPD and quit attempts), self-reported quit rates (7- and 30-day point prevalence abstinence [PPA]), and continuous abstinence rates (proportion who achieved uninterrupted abstinence) and duration. Exploratory regression analyses were performed to identify baseline characteristics associated with achievement of 7-day PPA, 30-day PPA, and continuous abstinence.

Results: A total of 319 participants completed onboarding (intention-to-treat [ITT]); 288/319 participants (90.3%) completed follow up (completers) at a mean of 7.2 (SD 1.2) months after onboarding. At final follow up, CPD were reduced by 52.6% (SE 2.1; $P<.001$) among all 319 participants, and most completers (152/288, 52.8%) reduced their CPD by at least 50%. Overall, most completers (232/288, 80.6%) made at least one quit attempt. Quit rates increased after the end of Pivot; using ITT analyses, 35.4% (113/319) achieved 7-day PPA and 31.3% (100/319) achieved 30-day PPA at final follow up compared with 32.0% (102/319) and 27.6% (88/319), respectively, at the end of the Pivot program. Continuous abstinence was achieved in about a quarter of those who onboarded (76/319, 23.8%) and in most who reported 30-day PPA at the end of Pivot (76/88, 86.4%), with a mean abstinence duration of 5.8 (SD 0.6) months. In exploratory regression analyses, lower baseline CPD, more positive baseline attitudes reflecting higher self-efficacy (higher confidence to quit and lower perceived difficulty of quitting), and higher education were associated with achieving abstinence.

Conclusions: This study provides the first longer-term outcomes of the Pivot smoking cessation program. At final follow up, quit rates increased and continuous abstinence was favorable; the majority who achieved abstinence at the end of Pivot sustained abstinence throughout follow up. Decreases in CPD persisted and most participants made a quit attempt. Overall, final follow-up

outcomes were stable or improved when compared to previous outcomes from the end of the program. These findings validate earlier results, and suggest that Pivot is an effective and durable solution for smoking cessation.

Trial Registration: ClinicalTrials.gov NCT03295643; https://clinicaltrials.gov/ct2/show/NCT03295643

(J Med Internet Res 2021;23(2):e25578) doi:10.2196/25578

KEYWORDS

smoking cessation; digital health; smartphone; digital sensor; carbon monoxide; breath sensor; biofeedback; mobile apps; health promotion; app

Introduction

Background

Tobacco use, primarily through cigarette smoking, is the leading cause of preventable disease, disability, and mortality in the United States [1]. Although smoking has declined over the last several decades, it remains a significant public health problem; in 2019, 14.0% of US adults (34.1 million people) were still current cigarette smokers [2].

Quitting smoking is one of the most important steps one can take for their health and can add as much as a decade to life expectancy; accordingly, most smokers (approximately 70%) want to quit [3,4]. Proven smoking cessation treatments include behavioral counseling and pharmacotherapy, which are now widely available. Use of these evidence-based approaches increases the rate of quitting by at least 40% [5-8]. However, use remains low, with less than one-third of smokers using any proven cessation treatments (eg, behavioral counseling, medication). As a result, most quit attempts are unassisted and more than 90% of these attempts are unsuccessful [4].

With room to expand and improve treatment options, the last decade has seen novel approaches to smoking cessation, including mobile and web-based options. A 2019 meta-analysis by Whittaker et al [9] assessed phone text messaging and app-based interventions for smoking cessation. In an assessment comprising 13 studies with 14,133 participants, the authors reported that automated text messaging interventions were more effective than minimal smoking cessation support (relative risk [RR] 1.54, 95% CI 1.19-2.00; I^2=71%). The authors also assessed five studies comprising 3079 participants, comparing a smoking cessation smartphone app with lower-intensity smoking cessation support (either a lower-intensity app or nonapp minimal support). This assessment provided no evidence that smartphone apps improved the likelihood of smoking cessation (RR 1.00, 95% CI 0.66-1.52; I^2 = 59%), but the authors noted that the evaluated evidence was of very low certainty due to inconsistency and imprecision, highlighting the need for more randomized controlled trials (RCTs) in this area. More recently, Bricker et al [10] performed an RCT comparing an acceptance and commitment therapy–based smoking cessation smartphone app (iCanQuit, n=1214) with a United States Clinical Practice Guideline (USCPG)-based app (QuitGuide, n=1201). At 12 months after randomization, iCanQuit participants had 1.49 times higher odds of quitting smoking compared with that of QuitGuide participants (28.2%, 293/1040 vs 21.1%, 225/1067; odds ratio [OR] 1.49, 95% CI 1.22-1.83; *P*<.001).

The Pivot program is a novel digital health intervention for smoking cessation that seeks to expand on these previous findings with respect to both intervention design and associated outcomes. Pivot comprises a multiphase mobile app, as well as the first Food and Drug Administration (FDA)-cleared personal carbon monoxide (CO) breath sensor, and dedicated human coaching delivered through in-app text messaging. Pivot is designed for individuals with varying levels of readiness to quit, and is based on the USCPG for treating tobacco use and dependence.

A prospective cohort study evaluated outcomes in 319 adult smokers who underwent the Pivot program (intention to treat [ITT] cohort); 272 (85.3%) participants completed the end-of-Pivot questionnaire (completer cohort) [11]. The study included individuals along the spectrum of readiness to quit; at study entry, the majority of participants (66.5%, 212/319) were not planning on quitting smoking in the next 30 days. Participant engagement, changes in attitudes toward quitting smoking, and changes in smoking behavior during and at the end of the Pivot program (mean 4.1, SD 1.4 months after enrollment) were assessed. Participants had a mean of 12.4 (SD 7.1) weeks of active program engagement, defined as at least one of the following per week: completing a breath sample; logging a cigarette; starting or completing a daily activity, challenge, or check-in; or messaging one's coach. Repeated-measures linear mixed-model analyses demonstrated positive changes in attitudes at the end of the prequit portion of the program, with increased confidence to quit (4.2 to 7.4, *P*<.001) and decreased expected difficulty in maintaining quit (3.1 to 6.8, *P*<.001). The quit attempt rate (ie, those making ≥1 quit attempt lasting ≥1 day) was 79.4% (216/272, completer analysis). At the end of Pivot, 7-day point prevalence abstinence (PPA) rates were 32.0% (102/319, ITT analysis) and 37.5% (102/272, completer analysis); 30-day PPA rates were 27.6% (88/319, ITT) and 32.4% (88/272, completer). Moreover, 30-day PPA rates were comparable among those ready and not ready to quit in the next 30 days at baseline. Of those not achieving abstinence, 25.9% (44/170, completer) achieved ≥50% reduction in cigarettes per day (CPD) at the end of the Pivot program.

Although these data are encouraging, there is an ongoing need to assess the durability of short-term results, and thereby establish longer-term outcomes in novel smoking cessation programs such as Pivot.

Objectives

This report focuses on participant outcomes 3 months after the completion of Pivot, including smoking behavior, quit rates,

continuous abstinence rates and durability, and predictors of abstinence.

Methods

Study Design

This was a prospective, open-label single-arm cohort study performed with institutional review board (IRB) approval. The study was performed remotely on an ambulatory basis. Study participants participated in the Pivot program and completed online study questionnaires. A detailed description of the study methodology was previously provided, with initial focus on outcomes at the end of active participation in Pivot [11].

Consent and Ethical Approval

All participants provided electronic informed consent before participation. The study was reviewed and approved by Solutions IRB (protocol number 2017/09/22) and was registered with Clinicaltrials.gov (NCT03295643).

Pivot Program

Pivot is a self-paced, comprehensive digital smoking cessation solution that includes an over-the-counter CO breath sensor, the multiphase Pivot mobile app, and human coaching delivered one-on-one through in-app text messaging [11].

Pivot Breath Sensor is a personal interactive FDA-cleared device that measures CO in exhaled breath. In line with wearable devices, the CO breath sensor provides real-time personal biometric data to users, enabling them to link their smoking behavior and CO values and track their progress in reducing or quitting smoking. This leverages the findings of several published studies [12-15] as well as expert opinion [16,17], which suggest that personal CO breath sample data can be educational and motivational, and may lead to changes in attitudes toward quitting and smoking behavior. To that end, the CO breath sensor is incorporated in the Pivot program as an engagement tool, with the intention that users will find their expired CO values informative and motivational.

In the multiphase Pivot app, participants could log cigarettes, follow trends in their CO values, complete educational and preparatory activities, set a quit date, make a quit plan, undertake short-term practice quits, learn about FDA-approved cessation medications, complete daily check-ins upon quitting smoking, and communicate with their coach.

Coaching was undertaken through asynchronous in-app text messaging, thus allowing participants to respond to coach-initiated contact or to initiate contact with their coach whenever it was convenient for them. Pivot coaches are trained specialists in tobacco cessation. The coach and Pivot participant are paired for the duration the participant is in Pivot to foster rapport and continuity. Coach-initiated contact included outreach 3 times a week from entry through the first 30 days after the quit date, once per week for the next 30 days, and then every other week for the last 30 days. Participants could initiate contact with their coach as frequently as desired.

The Pivot program's foundation is evidence-based and applies the USCPG-recommended "5 As" (Ask, Advise, Assess, Assist, and Arrange); tailors to one's readiness to quit smoking [18]; encourages the use of FDA-approved pharmacotherapy [18-21]; uses effective methods and supportive theories for smoking cessation (eg, motivational interviewing, cognitive behavioral therapy, and self-determination theory) [18,22-24]; and provides behavioral counseling through a live, dedicated coach [6,18,21,25].

Eligibility

To be eligible for participation, individuals had to meet all of the following eligibility criteria: 18-65 years of age, English-speaking, smoke ≥5 CPD, own and use a smartphone that is compatible with the Pivot app and breath sensor software (iPhone 5 and above, operating system iOS 9.0 and above, or Android operating system 4.4 and above), be employed for ≥20 hours a week, and live in the United States. Although we aim for broad availability of Pivot through multiple channels such as private and public insurers, direct-to-consumer, and not-for-profit foundations, at the time this study was performed, Pivot was initially only available to individuals through their employers (self-insured employers or employee wellness programs). As such, the employment requirement was applied to assess Pivot in individuals closely aligned with Pivot's initial user population.

Study Procedure

Study participants completed an online screening form, a screening phone call, electronic informed consent, web registration, and the baseline electronic questionnaire. They were mailed the breath sensor, which they set up independently using the labeling. Technical support was available as needed. Participants were assigned a coach with whom they worked for the duration of their participation in Pivot. Over the entire study, participants were compensated US $10 to $50 per completed study questionnaire and US $50 for returning the CO breath sensor for up to a total of US $315, using Visa gift cards. Specifically, for this follow-up portion of the study, participants were compensated US $50 for completing the final study questionnaire and US $50 for returning Pivot Breath Sensor if they had not yet done so. Compensation was not associated with use of the various components of Pivot, level of engagement, or smoking/quitting status.

Data Collection

Data were collected electronically through participant input in the Pivot online registration form, Pivot app, and online questionnaires. Study data were imported directly into a secure database (PostgreSQL, PostgreSQL Global Development Group).

Outcomes

Outcomes from this follow-up phase of the study focus on smoking behavior, quit rates, continuous abstinence rates and duration, and predictors of abstinence. For smoking behavior, outcomes include CPD and quit attempts. A quit attempt was defined as going at least 1 day without smoking cigarettes, even a single puff. Quit rates were self-reported and include 7- and 30-day PPA. Participants were considered to have achieved 7-day (30-day) PPA if they answered "no" to the following question: "In the last 7 (30) days have you smoked any

cigarettes, even a single puff?" As the Pivot program has no face-to-face contact, and data collection is achieved through remote means using the app and electronic questionnaires, biochemical verification of smoking status was not pursued in accordance with previous recommendations [26].

Although not listed as a preregistered outcome on Clinicaltrials.gov, we also evaluated continuous abstinence. The rationale for including continuous abstinence was to enhance the ability to compare outcomes in this study to those in other studies, to include an outcome that was a closer proxy for lifelong abstinence than PPA, and to include an outcome that was temporally closer to the Pivot program intervention than PPA [27]. Continuous abstinence includes the proportion of participants who achieved uninterrupted abstinence; a conservative definition was applied in which there was no grace period after the onset of abstinence, and any smoking (even a single puff) precluded designation of continuous abstinence. To be considered continuously abstinent, one had to report 30-day PPA on the end-of-Pivot questionnaire, 30-day PPA on the final questionnaire, and indicate a duration of abstinence that was equal to or greater than the number of days between the two questionnaires plus an additional 30 days. The average duration of abstinence is reported in those who achieved continuous abstinence. Finally, exploratory regression analyses were performed to identify the baseline characteristics associated with achievement of 7-day PPA, 30-day PPA, and continuous abstinence on the final questionnaire.

All participants were sent the end-of-Pivot and final follow-up questionnaires, regardless of their progress or completion status in the Pivot program.

Sample Size

Sample size was previously addressed [11]. A previous evaluation showed that attitudes toward quitting (motivation to quit, confidence to quit) are meaningful predictors of quit attempts [28]. On the basis of a previous assessment of 41 individuals using the first stage of Pivot (Explore), we estimated that the mean change in ratings assessing attitudes toward quitting (confidence to quit and expected difficulty maintaining quit) would be ≥1 (SD 4) just before reaching the Quit stage of the program [29]. On the basis of these estimates, there was 80% power to detect a significant difference in these ratings with a sample size of 101. As this was an initial study of the complete Pivot program, and in the context of known high attrition rates with mobile health apps [30-32], we applied conservative retention estimates drawn from other similar studies. Specifically, the target enrollment of 310 was estimated to yield at least 100 participants still engaged at the end of the Pivot program. The study enrolled 319 participants (ITT cohort) and 288 participants completed the final follow-up questionnaire (completer cohort).

Analyses

Changes in CPD were assessed from baseline to the final follow-up questionnaire. Participants served as their own controls, and comparisons were made to no change. To evaluate changes in CPD, repeated-measures linear mixed-model analyses were performed using a compound symmetric correlation matrix to model the repeated measures within participants. Because these measurements were taken at the same point in the study (not necessarily after the same amount of time, as progression through Pivot is self-paced), study stage (baseline vs final follow up) was used as a surrogate for time. To make specific comparisons across time, *F* statistics were computed using the results from the model.

Analyses were conducted to calculate the mean (SD) for normally distributed variables for actual data or mean (SE) for modeled data. Median (IQR) values were used in instances of non-normally distributed variables. A paired one-sample *t* test was used for numeric data. The Fisher exact or χ^2 test was used for comparisons of categorical data. The McNemar test was used for two-category match-paired data. Cohen κ statistic was used for three-category match-paired data.

In the assessment of quit rates (PPA), two sets of analyses were performed. In the ITT analysis, individuals who did not respond to PPA questions were assumed to be smoking. A study completer analysis was also performed, which only included individuals who completed the final follow-up questionnaire. Participants were sent the final follow-up questionnaire regardless of whether or not they completed the Pivot program. For additional assessments performed at the end of the study (quit attempts, proportion who reduced CPD by at least 50%), a study completer analysis was performed. This analysis approach comports with previous reports assessing app-based digital cessation programs [33,34].

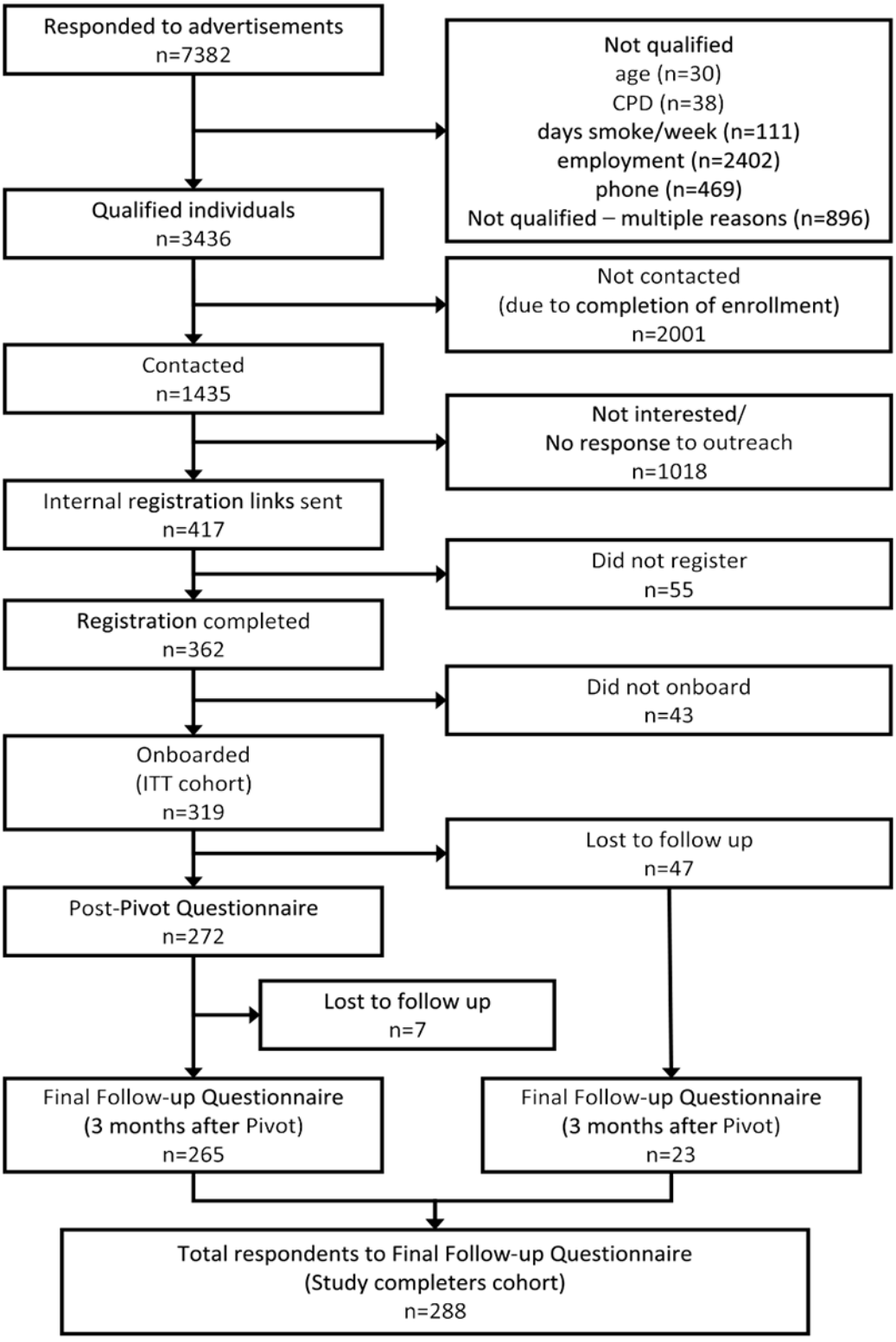
We performed exploratory post-hoc analyses using univariate logistic regression to explore associations between baseline characteristics and smoking behavior outcomes. We evaluated each independent baseline variable as a predictor in a separate model, with the binary outcomes of 7-day PPA, 30-day PPA, and continuous abstinence. We then performed multivariate logistic regression using forward selection of baseline variables with the binary outcomes of 7-day PPA, 30-day PPA, and continuous abstinence. Analyses were conducted using SAS Version 9.4 (SAS Institute, Cary, NC). Statistical significance was set at $P<.05$.

Results

Enrollment and Questionnaire Completion

A total of 319 participants completed onboarding and comprised the ITT cohort. At the end of active participation in Pivot, at a mean of 4.1 (SD 1.4) months after enrollment, 85.3% (272/319) of the participants completed the post-Pivot questionnaire [9]. At the end of the final follow-up period (3 months after completion of Pivot), at a mean of 7.2 (SD 1.2) months after enrollment, 90.3% (288/319) of participants completed the final follow-up questionnaire, who comprise the study completers cohort in this report. Study enrollment and attrition are depicted in the participant flow diagram shown in Figure 1.

Figure 1. Study participant flow: Consolidated Standards of Reporting Trials (CONSORT) diagram. ITT: intention to treat; CPD: cigarettes per day.



Baseline Characteristics

The study sample comprised 57.7% (184/319) women, had a mean age of 42.8 (SD 10.2) years, smoked a mean of 17.7 (SD 7.6) CPD at baseline, and had been smoking for a mean of 26.4 (SD 10.7) years. Participants represented 47 of the 50 US states; North Dakota, Nevada, and Arkansas were not represented. At baseline, 33.5% (107/319) of participants indicated that they were seriously thinking of quitting smoking in the next 30 days, 63.0% (201/319) indicated they were seriously thinking of quitting in the next 6 months, and 3.5% (11/319) indicated they were not thinking of quitting smoking. On average, participants had made 2.1 (SD 3.3) quit attempts over the past 12 months.

Smoking Behavior

Repeated-measures linear mixed-model analysis was performed with estimated final follow-up CPD values compared with baseline. There was an estimated 52.6% (SE 2.1) reduction in

CPD at final follow up (P<.001), which persisted from the end of Pivot. Table 1 details the CPD at baseline, the end of Pivot, and at final follow up.

Among those who completed the final follow-up assessment, most (152/288, 52.8%) reduced their CPD by ≥50%.

Focusing on the study completers who did not achieve at least 7-day PPA at final follow up (n=175), CPD decreased by 22.7% (SD 37.0), and 22.3% (39/175) reduced their CPD by ≥50%.

As reported previously, among the 170 participants who completed the end-of-Pivot questionnaire and did not achieve abstinence, 25.9% (44/170) achieved ≥50% reduction in CPD. Of these participants, 95.5% (42/44) completed final follow up, at which time 66.7% (28/42) reported 7-day PPA or ≥50% reduction in CPD. Specifically, 26.2% (11/42) achieved 7-day PPA (16.7%, 7/42 also achieved 30-day PPA) and 40.5% (17/42) did not achieve PPA but reported ≥50% reduction in CPD.

Overall, most completers (232/288, 80.6%) reported making at least one quit attempt during the study with an average of 2.9 (SD 3.7) quit attempts made per participant.

Table 1. Changes in cigarettes per day (CPD) based on the linear mixed model (N=319).

Time point	CPD		Change in CPD vs baseline		Percent change in CPD vs baseline	
	Mean (SE)	*P* value[a]	Mean (SE)	*P* value[a]	Mean (SE)	*P* value[a]
Baseline	17.7 (0.43)	N/A[b]	N/A	N/A	N/A	N/A
End of Pivot	8.0 (0.45)	<.001	–9.7 (0.42)	<.001	–54.5 (2.2)	<.001
Final follow up	8.5 (0.44)	<.001	–9.2 (0.41)	<.001	–52.6 (2.1)	<.001

[a]Compared to baseline.

[b]N/A: not applicable.

Quit Rates

Quit rates increased from the end of Pivot to final follow up. Specifically, at final follow up, 35.4% (113/319) achieved 7-day PPA and 31.3% (100/319) achieved 30-day PPA using ITT analysis. These rates increased from those obtained at the end of Pivot, when the 7-day PPA was 32.0% (102/319) and the 30-day PPA was 27.6% (88/319).

Similarly, at final follow up, 39.2% (113/288) achieved 7-day PPA and 34.7% (100/288) achieved 30-day PPA, using the study completer analysis. These rates increased from those obtained at the end of Pivot, when the 7-day PPA was 37.5% (102/272) and the 30-day PPA was 32.4% (88/272).

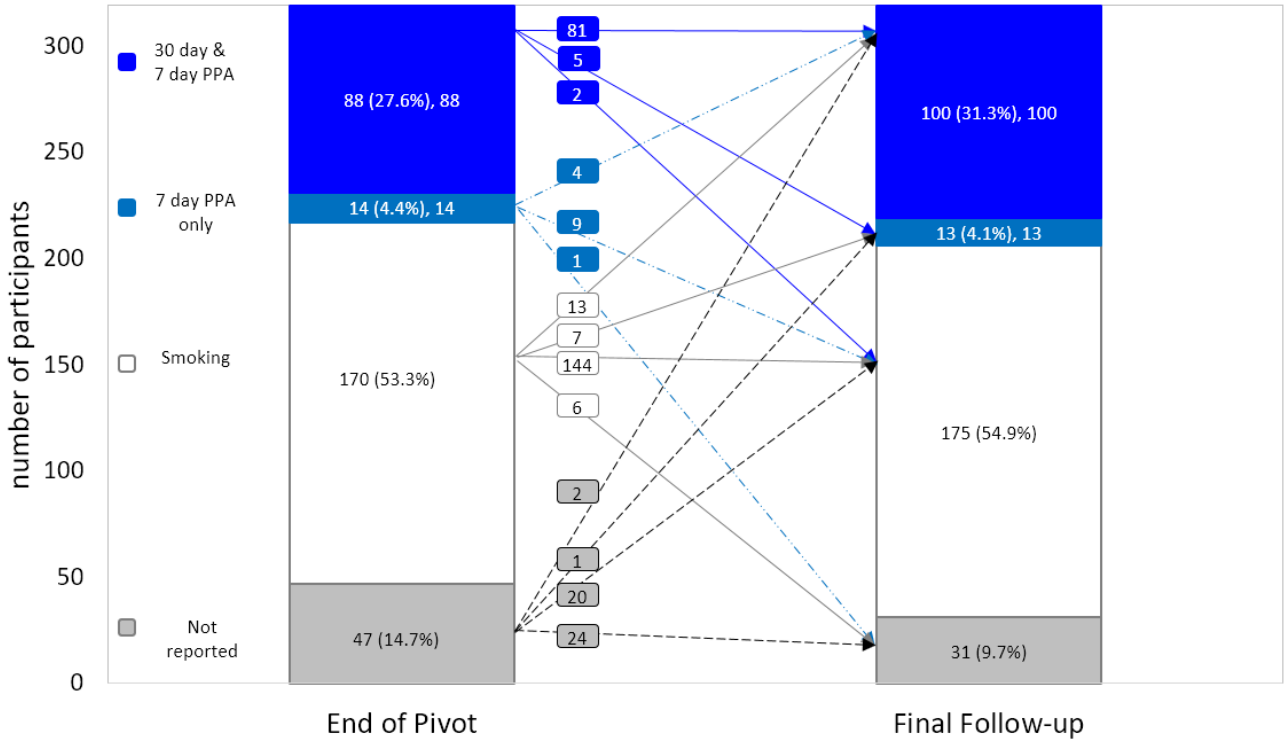
Assessing only the 265 participants who completed both the end-of-Pivot and final follow-up questionnaires, 41.5% (110/265) achieved 7-day PPA and 37.0% (98/265) achieved 30-day PPA. Among the 23 participants who did not complete the end-of-Pivot questionnaire but did complete the final follow up, 20 were still smoking, 1 reported 7-day PPA, and 2 reported 30-day PPA (as well as 7-day PPA) at final follow up (Figure 2).

Of the 88 participants who achieved 30-day PPA at the end of Pivot, 92.0% (81/88) reported 30-day PPA at final follow up.

From the end of Pivot to final follow up, there were 23 newly abstinent participants. Specifically, among the 217 participants who had not achieved at least 7-day PPA at the end of Pivot, 10.6% (23/217) reported abstinence at final follow up. Focusing on these 23 individuals, 15 achieved 30-day PPA (as well as 7-day PPA) and 8 achieved 7-day PPA (Figure 2).

Using all available data, 42.0% (134/319, ITT) achieved 7-day PPA and 35.4% (113/319, ITT) achieved 30-day PPA at some point during the study.

Figure 2. Participant smoking status at the end of Pivot and at final follow up.



Continuous Abstinence

Continuous abstinence was reported in 76 participants. Table 2 details continuous abstinence rates in various study subgroups.

Among all study participants, approximately a quarter achieved continuous abstinence. Among those who completed the final follow-up questionnaire, just over a quarter achieved continuous abstinence. Focusing on participants who made at least one quit attempt, nearly a third achieved continuous abstinence. Focusing on participants who reported 30-day PPA at the end of Pivot, the continuous abstinence rate was very high. The mean duration of continuous abstinence in this group was 5.8 (SD 0.6) months.

Table 2. Continuous abstinence rates in various study subgroups.

Study subgroup	N	Continuous abstinence rate, n (%)
Enrolled (intention to treat)	319	76 (23.8)
Completed final follow-up questionnaire (completers)	288	76 (26.4)
Made at least one quit attempt	232	76 (32.8)
Achieved 30-day point prevalence abstinence at the end of Pivot	88	76 (86.4)

Predictors of Abstinence

Focusing on all 319 participants who enrolled in the study, exploratory univariate regression analyses were performed to examine associations between participant baseline variables and final follow-up achievement of 7-PPA, 30-day PPA, and continuous abstinence; results are detailed in Table 3.

Lower baseline CPD, higher confidence to quit, and lower perceived difficulty maintaining quit were associated with an increased likelihood of achieving 7-day PPA, 30-day PPA, and continuous abstinence. Non-White Hispanic/Latino/Latina/Spanish origin was associated with a higher likelihood of achieving 7-day PPA, 30-day PPA, and continuous abstinence; however, this association should be considered with caution due to the low number of participants in this category (15/319, 4.7%).

Table 3. Univariate logistic regression analyses of baseline predictors of 7-day PPA, 30-day PPA, and continuous abstinence at final follow up among all study participants (N=319).

Baseline variable	N	7-day PPA[a]		30-day PPA		Continuous abstinence	
		OR[b] (95% CI)	*P* value[c]	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age	319	0.99 (0.97-1.01)	.33	0.98 (0.96- 1.00)	.08	0.98 (0.96-1.01)	.12
Gender							
Male	135	1 [Reference]	N/A[d]	1 [Reference]	N/A	1 [Reference]	N/A
Female	184	1.39 (0.87-2.23)	.17	1.30 (0.80-2.11)	.29	1.08 (0.64-1.83)	.76
Race/ethnicity							
White	264	1 [Reference]	N/A	1 [Reference]	N/A	1 [Reference]	N/A
American Indian/Alaska Native	4	0.66 (0.07-6.39)	.72	0.78 (0.08-7.62)	.83	1.21 (0.12-11.9)	.81
Asian	5	0.49 (0.05-4.46)	.53	0.59 (0.06-5.32)	.63	0.91 (0.10-8.28)	>.99
Black/African American	22	1.36 (0.56-3.31)	.50	1.09 (0.43-2.78)	.85	2.08 (0.83-5.19)	.20
Hispanic/Latino/Latina/Spanish origin	15	3.93 (1.31-11.9)	.02	3.51(1.21-10.20)	.02	3.18 (1.11-9.13)	.03
Native Hawaiian or other Pacific Islander	2	1.97 (0.12-31.8)	.63	2.34 (0.15-37.90)	.55	__[e]	—
Other race, ethnicity, or origin	7	0.79 (0.15-4.14)	.78	0.94 (0.18-4.90)	.94	1.45 (0.28-7.68)	.66
Education							
Bachelor's degree or greater	96	1 [Reference]	N/A	1 [Reference]	N/A	1 [Reference]	N/A
Less than bachelor's degree	223	1.60 (0.95-2.70)	.08	1.44 (0.84-2.45)	.18	1.52 (0.84-2.76)	.16
Income (US $)							
<50,000	136	1 [Reference]	N/A	1 [Reference]	N/A	1 [Reference]	NA
>50,000	173	0.79 (0.50-1.27)	.33	0.77 (0.47-1.25)	.29	1.21 (0.70-2.10)	.49
Did not answer	10	1.08 (0.29-4.00)	.91	0.81 (0.20-3.28)	.77	1.73 (0.42-7.14)	.45
Cigarettes per day	319	0.96 (0.93-0.99)	.01	0.97 (0.94-1.00)	.048	0.95 (0.91-0.99)	.01
Years smoking	319	0.99 (0.97-1.01)	.44	0.98 (0.96-1.01)	.18	0.98 (0.95-1.00)	.10
Quit attempts in last 12 months	319	1.03 (0.97-1.11)	.33	1.04 (0.97-1.11)	.32	1.04 (0.97-1.11)	.31
Stage of change							
Yes, within the next 30 days	107	1 [Reference]	N/A	1 [Reference]	N/A	1 [Reference]	NA
Yes, within the next 6 months	201	0.88 (0.54-1.43)	.59	0.96 (0.58-1.60)	.87	1.15 (0.66-2.00)	.78
No, not thinking of quitting	11	0.96 (0.26-3.48)	.95	1.23 (0.34-4.48)	.76	0.77 (0.16-3.80)	.35
Confidence to quit	319	1.11 (1.02-1.21)	.02	1.11 (1.02-1.21)	.02	1.17 (1.06-1.28)	.001
Perceived difficulty maintaining quit	319	1.14 (1.04-1.25)	.004	1.18 (1.08-1.30)	<.001	1.19 (1.08-1.31)	.003

[a]PPA: point prevalence abstinence.

[b]OR: odds ratio.

[c]*P* values are based on 95% Wald confidence limits.

[d]N/A: not applicable.

[e]—: No participants in this category.

Focusing specifically on the 88 participants who achieved 30-day PPA at the end of Pivot, exploratory univariate regression analyses were performed to examine associations between participant baseline variables and final follow-up achievement of continuous abstinence. None of the evaluated baseline variables was found to be associated with continuous abstinence.

In multivariate regression analyses, lower baseline CPD (OR 0.96, 95% CI 0.93-0.99; *P*=.02), lower perceived difficulty maintaining quit (OR 1.13, 95% CI 1.03-1.24; *P*=.01), and higher education (OR 1.72, 95% CI 1.01-2.94; *P*=.047) were associated with an increased likelihood of achieving 7-day PPA. Lower perceived difficulty maintaining quit (OR 1.18, 95% CI 1.08-1.30; *P*<.001) was associated with achieving 30-day PPA. Lower perceived difficulty maintaining quit (OR 1.13, 95% CI 1.02-1.26; *P*=.02) and higher confidence to quit (OR 1.12, 95% CI 1.01-1.24; *P*=.04) were associated with achieving continuous abstinence. Finally, no baseline variables were predictive of

continuous abstinence among the 88 individuals who achieved 30-day PPA at the end of Pivot.

Discussion

Principal Findings

The present report details longer-term follow-up outcomes from a prospective cohort study of 319 adult smokers who underwent the Pivot smoking cessation program. These outcomes, from a mean of 7.2 (SD 1.2) months postenrollment, focus on smoking behavior, quit rates via PPA, continuous abstinence rates and duration, and predictors of abstinence. To our knowledge, this is the first study assessing longer-term outcomes in a mobile smoking cessation program such as Pivot, which includes biofeedback via a personal portable CO breath sensor, education and guidance through a smartphone app, and support through text-based human coaching. At final follow up, CPD were reduced by approximately half, most participants had made a quit attempt, quit rates had increased from the end of the Pivot program, and approximately a quarter of participants achieved continuous abstinence. Regression analyses showed that lower CPD, attitudes toward quitting reflecting higher self-efficacy, and higher education level were associated with achieving abstinence. Overall, final follow-up outcomes persisted or improved when compared to previous outcomes from the end of the Pivot program.

Specific Findings

Scope for Comparison

Comparison with other studies is limited due to the novelty of digital smoking cessation programs that comprise a smartphone app, human-delivered text-based coaching, and a personal biofeedback device. In addition, among the few studies that did include smoking cessation interventions comparable to Pivot, differences in study design or population are significant enough to make comparison challenging. Taking these factors into consideration, we review the outcomes of our study with others below, with consideration of studies that employed a smoking cessation intervention similar to Pivot, expanding the assessment to include a broader group of digital smoking cessation interventions, and finally considering Pivot in the context of different types of smoking cessation interventions.

CPD Reduction

At final follow up, participants reduced CPD by 52.6%. Among those who did not achieve PPA, CPD were reduced by 22.7%, and 22.3% (39/175) reduced their CPD by ≥50%. Two additional studies have employed smoking cessation interventions similar to Pivot and included changes in CPD as an outcome. Webb et al [35] performed an RCT in adult smokers in the United Kingdom randomized to a digital therapeutic intervention (treatment, n=265) or very brief advice (control, n=265). The digital therapeutic intervention for smoking cessation comprised a smartphone app delivering cognitive behavioral therapy content, human coaching via phone and in-app chat, craving tools, and tracking capabilities. The control intervention was very brief advice applying the Ask, Advise, Act model. Half of the participants received a personal CO breath sensor that was used to measure their exhaled CO and validate self-reported abstinence. Eligibility criteria included readiness to quit in the next 30 days. Participants had an in-person baseline visit, and then self-reported outcomes via phone or online at 4 weeks after the quit date. Participants set a quit date an average 16 days postrandomization. All participants were offered free nicotine replacement therapy (NRT) for 3 months, which was used by 59.1% (133/225, treatment) and 63.2% (146/231, control) of participants (risk ratio 0.94, 95% CI 0.81-1.08). At 4 weeks post quit date, mean CPD were reduced in those who failed to achieve abstinence by 48.1% in the treatment group and by 48.9% in the control group. The provision of NRT that was used by most participants, the 1-month endpoint that occurred during the active treatment phase of the intervention, and inclusion only of individuals ready to quit smoking in the next 30 days in the Webb study are notable study design differences that likely contribute to differences in CPD reduction from those found in the present study.

The second study that evaluated a smoking cessation intervention similar to Pivot, performed by Krishnan et al [36], was an RCT in which adult smokers in the United States received brief advice along with a personal CO breath sensor and the COach2Quit app (intervention, n=50) or brief advice only (control, n=52). The Coach2Quit app prompted the user to set a quit date, provided reminders to complete two breath samples a day, sent response messages from a text message library to users after breath samples were provided based on their CO results, and provided graphical representation of user CO readings. Eligibility criteria included willingness to set a quit date within 2 weeks of the baseline assessment. Follow-up visits were conducted at 14 days and 30 days from baseline. The median CPD among all participants decreased from 10 at baseline to 5 in the intervention group and to 6 in the control group at 30 days, for an approximate reduction of 40%-50%. This is in range with the 52.6% CPD reduction for all participants found in our study; however, it is unknown if and how CPD reduction changed in the Krishnan study after the active intervention phase.

Looking more broadly at digital health interventions, Garrison et al [37] recently reported results from an RCT comparing the efficacy of 22 days of mobile mindfulness training through the Craving to Quit app with app-based experience sampling (n=143) versus 22 days of app-based experience sampling only (n=182). At 6 months, CPD were reduced but not different between the two groups; the app group reduced CPD by 43.8% (from 16.0, SD 7.1 to 9.0, SD 7.8) and the experience sampling–only group reduced CPD by 45.7% (from 16.2, SD 8.2 to 8.8, SD 9.0). Finally, Alessi et al [38] performed an RCT with 90 participants randomized to usual care and ecological monitoring with abstinence reinforcement (mobile health reinforcement) or without reinforcement (mobile health monitoring). Usual care was 8 weeks of transdermal nicotine and twice-weekly telephone counseling. Ecological monitoring was administered through an interactive voice response system that prompted participants to conduct 1-3 CO breath tests daily, video record the process, and submit the videos. Participants in the abstinence reinforcement group could earn prizes for on-time CO breath tests with CO values consistent with abstinence. At 6 months, CPD were reduced by 47% (baseline 17.6 to 9.3 CPD

at 6 months) in the mobile health reinforcement group and by 55% (baseline 20.0 to 9.0 at 6 months) in the mobile health monitoring group.

Overall, assessing the present and aforementioned studies, digital smoking cessation interventions have reported CPD reduction by about half at 6 months. Studies that have assessed interventions similar to Pivot have reported similar reductions achieved earlier, at 1 month; however, the durability of the reduction in these instances is unknown.

PPA Rates

PPA between the end of Pivot and final follow up 3 months later increased; using ITT analyses, 35.4% (113/319) achieved 7-day PPA and 31.3% (100/319) achieved 30-day PPA at final follow up, compared to 32.0% (102/319) and 27.6% (88/319), respectively, at the end of the Pivot program.

Comparing our 35.4% final follow-up 7-day PPA rate to other studies with interventions similar to Pivot, Webb et al [35] reported a 44.5% 7-day PPA rate at 1 month. The approximate 35%-45% 7-day PPA range seems within reason when considering that the Webb study included the provision of NRT and required participants to be ready to quit smoking within the next 30 days at enrollment for study eligibility, acknowledging that longer-term data from the Webb study will be informative. Masaki et al [39] performed an RCT of adult smokers recruited from smoking cessation clinics in Japan. Participants were randomized to the intervention, which included a 12-week standard smoking cessation treatment plus the CureApp Smoking Cessation (CASC) system (n=285), or to the control, which consisted of the 12-week standard smoking cessation treatment plus a control app (n=287). The CASC system comprised a smartphone app, paired mobile exhaled CO breath sensor, connected cloud system to upload data, and web-based PC software for physicians. The 12-week standard smoking cessation treatment included five in-person visits with counseling and physician-provided pharmacotherapy of varenicline or nicotine patch. The CASC system and control app were used for 24 weeks. Eligibility criteria included intention to quit smoking immediately. Seven-day PPA at 24 weeks was achieved in 72.3% of the CASC intervention group participants and in 58.2% of the control group participants ($P<.01$). Considering the 58.2% 7-day PPA rate in the control arm, this high 7-day PPA rate in the CASC intervention arm likely reflects what a program similar to Pivot adds when used as a supplement to a traditional intensive smoking cessation program in individuals ready to quit smoking.

Looking more broadly at outcomes from digital smoking cessation interventions, 7-day PPA at 6 months range from 9.8% to 29.6%, with most falling between 17% and 25% [10,37,38,40,41]. Our higher 7-day PPA rate of 35.4% aligns with expectations considering Pivot's additional features of the personal breath sensor and coaching.

Focusing on 30-day PPA, 31.3% (100/319) achieved this outcome at final follow up in our study. We did not find comparable data in studies that included interventions similar to Pivot. Broadening the scope to include additional digital smoking cessation interventions, 6-month 30-day PPA outcomes ranged from 12.9% to 21.8%, with an additional study reporting a 26.2% 30-day PPA rate at 8 weeks [10,34,40]. Similar to our expectation, we believe that the additional features of Pivot beyond the app likely contributed to the higher 30-day PPA rate.

Continuous Abstinence

Continuous abstinence at final follow up was reported in about a quarter of all participants (76/319, 23.8%) and in about a third of those who made a quit attempt (32.8%, 76/232) in our study. Masaki et al [39] reported that 63.9% of participants achieved continuous abstinence at 6 months. These differences in continuous abstinence likely reflect outcomes when Pivot is used as the sole intervention among smokers who represent the entire spectrum of readiness to quit in contrast to when a Pivot-like program is used as an adjunct to an intensive smoking cessation program among individuals ready to quit smoking. Expanding the scope of studies to the broader category of digital smoking cessation interventions, the rates for continuous abstinence at 6 months range from 4.8% to 19.8%, with most falling between 10% and 16% [37,38,40,42]. Overall, there appears to be a trend reflecting the intensity and comprehensiveness of the intervention, with lower-intensity programs reporting continuous abstinence at 6 months in about 10%-16% of participants, mid-intensity programs such as Pivot reporting continuous abstinence in about a quarter of all participants or a third of participants who make a quit attempt, and high-intensity programs achieving continuous abstinence in over half of participants.

Considering Pivot in the context of different types of smoking cessation interventions, we turn to meta-analyses to provide insight on longer-term outcomes. Whittaker et al [9] published a meta-analysis in 2019 assessing automated mobile phone text messaging and app-based interventions for smoking cessation. Long-term abstinence (defined as smoking cessation at 6 months or longer using the most stringent measure available) was higher in text messaging–based interventions compared to minimal smoking cessation support (13 studies, 14,133 participants; RR 1.54, 95% CI 1.19-2.00; I^2=71%). Evaluating both high- and low-intensity text message–based interventions using data pooled from three studies, the authors reported average ≥6-month abstinence rates of 26.6%-27.1%. A similar effect was not seen for comparison of smartphone app interventions to lower-intensity smoking cessation support (5 studies, 3079 participants; RR 1.00, 95% CI 0.66-1.52; I^2=59%); however, the authors noted the need for additional data to further assess these interventions, as this particular evaluation comprised 5 studies with many additional studies ongoing at the time of publication.

Matkin et al [43] published a meta-analysis in 2019 on telephone counseling for smoking cessation. In studies that recruited smokers who underwent proactive telephone counseling (ie, counseling that was not delivered through calling a helpline), the counseling increased quit rates (RR 1.25, 95% CI 1.15-1.35; I^2=52%; 65 trials, 41,233 participants). The authors reported that based on a control group quit rate of 11%, telephone counseling would produce an absolute increase of 2%-4%, resulting in a ≥6-month quit rate of 13%-15%. In a 2017

meta-analysis, Lancaster et al [6] reported that individual counseling increases the likelihood of cessation compared with less intensive support. Based on pooling 27 trials comprising 11,100 participants, individual counseling, when used independently of pharmacotherapy, was estimated to increase cessation by 40% to 80% after at least 6 months. Assuming a control group quit rate of 7% from a brief intervention, individual counseling would be expected to result in an absolute increase of 3%-5%, yielding a 10%-12% quit rate. Finally, in a 2019 meta-analysis, Hartmann-Boyce et al [5] reported that NRT increased quit rates compared to control, by an amount that depended on the baseline quit rate. For example, for an expected quit rate of 3%-5% in people attempting to quit on their own, NRT might increase the quit rate by 2%-3%. However, if the expected quit rate of a population was 15%, another 8% might be expected to quit with NRT use.

Summarizing these findings from meta-analyses, long-term quit rates for various types of smoking cessation interventions include 10%-15% for telephone or individual counseling alone, 26%-27% for automated text messaging, and 5% to more than 23% for NRT alone (depending on the expected quit rate in the population at baseline). Acknowledging that consideration of these data is primarily to establish context for the outcomes of Pivot, and that direct comparison of our cohort study with large meta-analyses is not appropriate, we believe that our 23.8% continuous abstinence rate in all participants is reasonable and encouraging. As a program that includes established evidence-based components such as coaching, as well as novel aspects that aim to leverage nascent but promising approaches to smoking cessation such as biofeedback via a personal breath sensor, we would expect to improve upon existing interventions. Certainly, this should be born out with future additional investigation and data.

Abstinence Duration

Hughes et al [44] established the significance of abstinence duration, reporting that abstinence stabilizes at about 6 months. Zhou et al [45] reported similar findings in their evaluation of 2431 smokers who intended to stop smoking in the next 3 months. They followed these individuals periodically for 18 months via internet questionnaires; after 6 months of abstinence, the relapse rate dropped below 20% and the cumulative relapse rate reached a plateau. Herd et al [46] detailed similar findings in a longitudinal survey of 1296 ex-smokers in the general population who quit on their own (not in an interventional study), reporting that a duration of abstinence of 31-182 days was associated with a 58% continuous abstinence rate, whereas duration of abstinence of 183-365 days was associated with a 78% continuous abstinence rate. These data bring further confidence to our results; with a continuous abstinence duration of approximately 6 months (mean 5.8, SD 0.6 months), there is reason to expect stability in our continuous abstinence rate.

Approaching continuous abstinence from a different angle, in a systematic review, Hughes et al [27] assessed the relationship between PPA rates and prolonged abstinence in studies with point prevalence durations of up to 7 days and follow ups of at least 6 months from the quit date. They reported that point prevalence and prolonged abstinence were highly correlated (r=0.88) and that prolonged abstinence averaged 0.74 that of PPA, indicating that approximately three-quarters of those who achieve point prevalence will achieve prolonged abstinence. In the most comparable analysis from our data, 79.4% (81/102) of study participants who achieved 7-day PPA at the end of Pivot achieved continuous abstinence of approximately 6 months duration (mean 5.6, SD 0.7 months), findings which align with those reported by Hughes et al [27].

Predictors of Abstinence

Exploratory univariate regression modeling demonstrated that among all study participants, lower baseline CPD, higher self-efficacy through confidence to quit, and lower perceived difficulty of maintaining quit were associated with achieving 7-day PPA, 30-day PPA, and continuous abstinence at final follow up, approximately 7 months from enrollment. Multivariate regression modeling found that lower perceived difficulty of maintaining quit was associated with 7- and 30-day PPA and continuous abstinence, lower CPD and higher education were associated with 7-day PPA, and higher confidence to quit was associated with continuous abstinence.

These results are consistent with previous findings. In the aforementioned study of 1296 ex-smokers by Herd et al [46], relapse was associated with lower abstinence self-efficacy. Smit et al [47] assessed predictors of successful quit attempts among 570 smokers motivated to quit in the next 6 months who were randomized to the control group in a web-based smoking cessation intervention study. They reported that self-efficacy was the main factor in predicting quit attempt success.

Nicotine dependence has also been shown to be a predictor of successful cessation [26]. In a systematic literature review of adult general population smokers, Vangeli et al [28] reported that cigarette dependence consistently predicted success after a quit attempt. Hymnowitz et al [48] performed a cohort tracking telephone survey in 13,415 smokers over 5 years. They reported that predictors with the largest RR values for smoking cessation were those associated with nicotine dependence, including CPD. Thereby, it is not surprising that in our study, lower baseline CPD was associated with achieving abstinence.

Finally, higher education has been shown to be a predictor for success in cessation. In a study including 887 smokers undergoing a smoking cessation program through their workplace, a higher educational level (OR 1.81, 95% CI 1.06-3.09, P=.03) predicted successful cessation [49]. In a study including 4397 smokers who participated in a two-armed RCT assessing computer-tailored smoking cessation advice in the United Kingdom, Kale et al [50] reported that a higher reading level was associated with successful quitting (OR 1.62, 95% CI 1.19-2.21). Reid et al [51] evaluated smokers in Canada, the United Kingdom, Australia, and the United States from the first five waves (2002-2006/2007) of the International Tobacco Control Four Country Survey (35,532 observations from 16,458 respondents). They reported that smokers with a high education level were more likely to be abstinent for 1 and 6 months (OR 1.20, 95% CI 1.00-1.44 and OR 1.30, 95% CI 1.05-1.62, respectively).

Limitations

Limitations of this study were discussed previously [11] and include limited representation of individuals who use Android smartphones, are employed less than 20 hours per week, and are not seriously thinking of quitting smoking. Regarding the smartphone platform, Ubhi et al [52] reported differing behavior among users of a smoking cessation app (SmokeFree28 app) between Android and iOS users, with iOS users being more likely to have made a quit attempt in the last 12 months and set their quit date on the day of registration, and Android users being more likely to have used smoking cessation medication in their quit attempt. Baseline intention to quit (Stage of Change) as well as factors associated with socioeconomic status have been documented as predictors for quit attempts and success in quitting smoking [46,47,49-51]. Collectively, these findings highlight the need for additional data on the Pivot program in members of these groups with limited representation in this study.

In addition, although multiple efforts were taken to minimize the impact of participant study payments, including keeping individual payments under US $50, incorporating a several-week delay between questionnaire completion and payment receipt, and not linking payment to use of program components or smoking outcomes, we cannot exclude some influence of study payment on participant behavior.

Owing to the sequential nature of Pivot, we elected to obtain outcome data as participants advanced through the self-paced program. The final follow-up data in our study is from a mean 7.2 (SD 1.2) months postenrollment, a time point that we believe was reasonable to consider with 6-month outcomes from other studies. This approach is different from the more traditional 30-, 90-, and 180-day assessments that are linked directly to enrollment date, and this difference is worth acknowledging as it limits direct comparison between studies.

The self-reported nature of smoking status is also a possible limitation. Biochemical verification of results was not sought for a few reasons. First, this study comprised general population smokers and was conducted entirely remotely with all data collection performed electronically via online questionnaires and through the Pivot app. Based on these study characteristics, biochemical verification was not pursued in accordance with previous recommendations [26,53]. Second, the breath sensor was employed as an educational and motivational tool in this study. We were concerned that implementing verification of smoking status with the breath sensor would instill a sense of policing that might detract from participant perception and experience of the sensor. Finally, while we acknowledge the possible limitation of over-reporting cessation with self-reporting smoking status, the literature suggests that the rate of occurrence is less than 10% in general population smokers. Specifically, Gorber et al [54] performed a systematic review of 54 studies to measure the concordance between self-reported smoking status and smoking status biochemically verified through measures of cotinine. The mean difference between self-reported and measured prevalence estimates was −4.8% for studies that measured cotinine in saliva; −6.2% for those measuring cotinine in serum, blood, or plasma; and −9.4% when cotinine was measured in urine.

Finally, although participants were provided education about NRT via the Pivot app and coaching in the study, they were not provided NRT. Given the well-established positive impact of NRT on smoking cessation [5,18-21], inclusion of NRT in Pivot would have likely further increased quit rates. The provision of NRT has since been added to the Pivot program; accordingly, future studies are warranted.

Conclusions

This follow-up study provides the first longer-term outcomes of Pivot, an inclusive and innovative smoking cessation program that employs a smartphone app, biofeedback through a personal portable CO breath sensor, and in-app text-based coaching. At final follow up, approximately 7 months after enrollment, quit rates increased and continuous abstinence was favorable. Most participants made a quit attempt. The emergence of newly abstinent participants and persistent decreases in CPD 3 months after active program participation underscore the sustained learning and impact from Pivot. These results validate earlier findings, and suggest that Pivot is an effective and durable solution for smoking cessation.

Acknowledgments

The authors thank Heather Patrick, PhD, for her critical contributions to the study design, early data analysis, and drafting of the previous manuscript. We also thank Lydia Tesfamariam for assistance with recruiting, collaborative tracking of participants, study payments, and database management and analysis. The work of JG is supported, in part, by National Institutes of Health Grant P30 DK034987.

Authors' Contributions

JM, DU, and CF designed the study. JM and CF recruited participants. JM ran the study. CF managed the database and performed data analyses. JG performed the statistical analyses. JM prepared the original draft of the manuscript. CF, DB, DU, and JG reviewed and edited the manuscript before submission.

Conflicts of Interest

JM, CF, DB, and DU are current employees of Carrot Inc, the developer of the app and devices used in this study. They receive salary and stock options from Carrot Inc. DU is the President and CEO of Carrot Inc and an investor in the company.

References

1. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The Health Consequences of Smoking—50 Years of Progress A Report of the Surgeon General. Atlanta, GA: Centers for Disease Control and Prevention; 2014. URL: https://www.ncbi.nlm.nih.gov/books/NBK179276/ [accessed 2021-02-06]
2. Cornelius ME, Wang TW, Jamal A, Loretan CG, Neff LJ. Tobacco Product Use Among Adults - United States, 2019. MMWR Morb Mortal Wkly Rep 2020 Nov 20;69(46):1736-1742. [doi: 10.15585/mmwr.mm6946a4] [Medline: 33211681]
3. Jha P, Ramasundarahettige C, Landsman V, Rostron B, Thun M, Anderson RN, et al. 21st-century hazards of smoking and benefits of cessation in the United States. N Engl J Med 2013 Jan 24;368(4):341-350. [doi: 10.1056/NEJMsa1211128] [Medline: 23343063]
4. Babb S, Malarcher A, Schauer G, Asman K, Jamal A. Quitting Smoking Among Adults - United States, 2000-2015. MMWR Morb Mortal Wkly Rep 2017 Jan 06;65(52):1457-1464. [doi: 10.15585/mmwr.mm6552a1] [Medline: 28056007]
5. Hartmann-Boyce J, Chepkin SC, Ye W, Bullen C, Lancaster T. Nicotine replacement therapy versus control for smoking cessation. Cochrane Database Syst Rev 2018 May 31;5:CD000146 [FREE Full text] [doi: 10.1002/14651858.CD000146.pub5] [Medline: 29852054]
6. Lancaster T, Stead LF. Individual behavioural counselling for smoking cessation. Cochrane Database Syst Rev 2017 Mar 31;3:CD001292 [FREE Full text] [doi: 10.1002/14651858.CD001292.pub3] [Medline: 28361496]
7. Stead LF, Carroll AJ, Lancaster T. Group behaviour therapy programmes for smoking cessation. Cochrane Database Syst Rev 2017 Mar 31;3:CD001007 [FREE Full text] [doi: 10.1002/14651858.CD001007.pub3] [Medline: 28361497]
8. Stead LF, Koilpillai P, Fanshawe TR, Lancaster T. Combined pharmacotherapy and behavioural interventions for smoking cessation. Cochrane Database Syst Rev 2016 Mar 24;3:CD008286. [doi: 10.1002/14651858.CD008286.pub3] [Medline: 27009521]
9. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
10. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of Smartphone Applications for Smoking Cessation: A Randomized Clinical Trial. JAMA Intern Med 2020 Sep 21;180(11):1-9. [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
11. Marler JD, Fujii CA, Utley DS, Tesfamariam LJ, Galanko JA, Patrick H. Initial assessment of a comprehensive digital smoking cessation program that incorporates a mobile app, breath sensor, and coaching: cohort study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e12609 [FREE Full text] [doi: 10.2196/12609] [Medline: 30670372]
12. Beard E, West R. Pilot study of the use of personal carbon monoxide monitoring to achieve radical smoking reduction. J Smok Cessat 2012 Jul 13;7(1):12-17. [doi: 10.1017/jsc.2012.1]
13. Choi W, Kim C, Lee O. Effects of brief smoking cessation education with expiratory carbon monoxide measurement on level of motivation to quit smoking. Korean J Fam Med 2013 May;34(3):190-198 [FREE Full text] [doi: 10.4082/kjfm.2013.34.3.190] [Medline: 23730486]
14. Risser NL, Belcher DW. Adding spirometry, carbon monoxide, and pulmonary symptom results to smoking cessation counseling: a randomized trial. J Gen Intern Med 1990;5(1):16-22. [doi: 10.1007/BF02602303] [Medline: 2405112]
15. Jamrozik K, Vessey M, Fowler G, Wald N, Parker G, Van Vunakis H. Controlled trial of three different antismoking interventions in general practice. Br Med J (Clin Res Ed) 1984 May 19;288(6429):1499-1503 [FREE Full text] [doi: 10.1136/bmj.288.6429.1499] [Medline: 6426618]
16. Fagerström KO. Assessment of the smoker who wants to quit. Monaldi Arch Chest Dis 2001 Apr;56(2):124-127. [Medline: 11499300]
17. Bittoun R. Carbon monoxide meter: the essential clinical tool — the 'stethoscope' — of smoking cessation. J Smok Cessat 2012 Feb 21;3(2):69-70. [doi: 10.1375/jsc.3.2.69]
18. Tobacco Use and Dependence Guideline Panel. Treating Tobacco Use Dependence Update. Agency for Healthcare Research and Quality. Rockville, MD: US Department of Health and Human Services; 2008 May. URL: https://www.ahrq.gov/prevention/guidelines/tobacco/index.html [accessed 2021-02-06]
19. Coverage for tobacco use cessation treatments. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/tobacco/quit_smoking/cessation/coverage/ [accessed 2020-10-27]
20. Stead LF, Perera R, Bullen C, Mant D, Hartmann-Boyce J, Cahill K, et al. Nicotine replacement therapy for smoking cessation. Cochrane Database Syst Rev 2012 Nov 14;11:CD000146. [doi: 10.1002/14651858.CD000146.pub4] [Medline: 23152200]
21. Final Recommendation Statement - Tobacco Smoking Cessation in Adults, Including Pregnant Women: Behavioral and Pharmacotherapy Interventions. US Preventive Services Task Force. 2018. URL: https://www.uspreventiveservicestaskforce.org/uspstf/document/RecommendationStatementFinal/tobacco-use-in-adults-and-pregnant-women-counseling-and-interventions [accessed 2020-10-27]
22. Miller W, Rollnick S. Motivational Interviewing: Preparing People For Change Second Edition. New York City: Guilford Press; 2002.

23. Lindson-Hawley N, Thompson TP, Begh R. Motivational interviewing for smoking cessation. Cochrane Database Syst Rev 2015 Mar 02(3):CD006936. [doi: 10.1002/14651858.CD006936.pub3] [Medline: 25726920]
24. Williams GC, Niemiec CP, Patrick H, Ryan RM, Deci EL. The importance of supporting autonomy and perceived competence in facilitating long-term tobacco abstinence. Ann Behav Med 2009 Jun;37(3):315-324 [FREE Full text] [doi: 10.1007/s12160-009-9090-y] [Medline: 19373517]
25. Abrams D, Niaura R, Brown R, Emmons K, Goldstein M, Monti P. The Tobacco Dependence Treatment Handbook: A Guide To Best Practices. New York City: Guilford Press; 2003.
26. SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. Nicotine Tob Res 2002 May;4(2):149-159. [doi: 10.1080/14622200210123581] [Medline: 12028847]
27. Hughes JR, Carpenter MJ, Naud S. Do point prevalence and prolonged abstinence measures produce similar results in smoking cessation studies? A systematic review. Nicotine Tob Res 2010 Jul;12(7):756-762 [FREE Full text] [doi: 10.1093/ntr/ntq078] [Medline: 20504946]
28. Vangeli E, Stapleton J, Smit ES, Borland R, West R. Predictors of attempts to stop smoking and their success in adult general population samples: a systematic review. Addiction 2011 Dec;106(12):2110-2121. [doi: 10.1111/j.1360-0443.2011.03565.x] [Medline: 21752135]
29. Patrick H, Fujii CA, Glaser DB, Utley DS, Marler JD. A comprehensive digital program for smoking cessation: assessing feasibility in a single-group cohort study. JMIR Mhealth Uhealth 2018 Dec 18;6(12):e11708 [FREE Full text] [doi: 10.2196/11708] [Medline: 30563807]
30. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015 Jan 16;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
31. Perro J. Mobile apps: what's a good retention rate?. 2018. URL: https://uplandsoftware.com/localytics/resources/blog/mobile-apps-whats-a-good-retention-rate/ [accessed 2020-12-30]
32. Clement J. Retention rate on day 1 and day 30 of mobile app installs worldwide as of August 2020, by category. Statista. URL: https://www.statista.com/statistics/259329/ios-and-android-app-user-retention-rate/ [accessed 2020-12-30]
33. Bricker JB, Copeland W, Mull KE, Zeng EY, Watson NL, Akioka KJ, et al. Single-arm trial of the second version of an acceptance & commitment therapy smartphone application for smoking cessation. Drug Alcohol Depend 2017 Jan 01;170:37-42 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.10.029] [Medline: 27870987]
34. Iacoviello BM, Steinerman JR, Klein DB, Silver TL, Berger AG, Luo SX, et al. Clickotine, a personalized smartphone app for smoking cessation: initial evaluation. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e56 [FREE Full text] [doi: 10.2196/mhealth.7226] [Medline: 28442453]
35. Webb J, Peerbux S, Smittenaar P, Siddiqui S, Sherwani Y, Ahmed M, et al. Preliminary outcomes of a digital therapeutic intervention for smoking cessation in adult smokers: randomized controlled trial. JMIR Ment Health 2020 Oct 06;7(10):e22833 [FREE Full text] [doi: 10.2196/22833] [Medline: 33021488]
36. Krishnan N, Elf JL, Chon S, Golub JE. COach2Quit: a pilot randomized controlled trial of a personal carbon monoxide monitor for smoking cessation. Nicotine Tob Res 2019 Oct 26;21(11):1573-1577 [FREE Full text] [doi: 10.1093/ntr/nty182] [Medline: 30169740]
37. Garrison KA, Pal P, O'Malley SS, Pittman BP, Gueorguieva R, Rojiani R, et al. Craving to Quit: a randomized controlled trial of smartphone app-based mindfulness training for smoking cessation. Nicotine Tob Res 2020 Mar 16;22(3):324-331 [FREE Full text] [doi: 10.1093/ntr/nty126] [Medline: 29917096]
38. Alessi SM, Rash CJ, Petry NM. A randomized trial of adjunct mHealth abstinence reinforcement with transdermal nicotine and counseling for smoking cessation. Nicotine Tob Res 2017 Mar 01;19(3):290-298 [FREE Full text] [doi: 10.1093/ntr/ntw155] [Medline: 27613901]
39. Masaki K, Tateno H, Nomura A, Muto T, Suzuki S, Satake K, et al. A randomized controlled trial of a smoking cessation smartphone application with a carbon monoxide checker. NPJ Digit Med 2020;3:35. [doi: 10.1038/s41746-020-0243-5] [Medline: 32195370]
40. Baskerville NB, Struik LL, Guindon GE, Norman CD, Whittaker R, Burns C, et al. Effect of a mobile phone intervention on quitting smoking in a young adult population of smokers: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e10893 [FREE Full text] [doi: 10.2196/10893] [Medline: 30355563]
41. Danaher BG, Tyler MS, Crowley RC, Brendryen H, Seeley JR. Outcomes and device usage for fully automated internet interventions designed for a smartphone or personal computer: The MobileQuit smoking cessation randomized controlled trial. J Med Internet Res 2019 Jun 06;21(6):e13290 [FREE Full text] [doi: 10.2196/13290] [Medline: 31172967]
42. BinDhim NF, McGeechan K, Trevena L. Smartphone Smoking Cessation Application (SSC App) trial: a multicountry double-blind automated randomised controlled trial of a smoking cessation decision-aid 'app'. BMJ Open 2018 Jan 21;8(1):e017105 [FREE Full text] [doi: 10.1136/bmjopen-2017-017105] [Medline: 29358418]
43. Matkin W, Ordóñez-Mena JM, Hartmann-Boyce J. Telephone counselling for smoking cessation. Cochrane Database Syst Rev 2019 May 02;5:CD002850 [FREE Full text] [doi: 10.1002/14651858.CD002850.pub4] [Medline: 31045250]
44. Hughes JR, Keely J, Naud S. Shape of the relapse curve and long-term abstinence among untreated smokers. Addiction 2004 Jan;99(1):29-38. [doi: 10.1111/j.1360-0443.2004.00540.x] [Medline: 14678060]

45. Zhou X, Nonnemaker J, Sherrill B, Gilsenan AW, Coste F, West R. Attempts to quit smoking and relapse: factors associated with success or failure from the ATTEMPT cohort study. Addict Behav 2009 Apr;34(4):365-373. [doi: 10.1016/j.addbeh.2008.11.013] [Medline: 19097706]
46. Herd N, Borland R, Hyland A. Predictors of smoking relapse by duration of abstinence: findings from the International Tobacco Control (ITC) Four Country Survey. Addiction 2009 Dec;104(12):2088-2099 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02732.x] [Medline: 19922574]
47. Smit ES, Hoving C, Schelleman-Offermans K, West R, de Vries H. Predictors of successful and unsuccessful quit attempts among smokers motivated to quit. Addict Behav 2014 Sep;39(9):1318-1324. [doi: 10.1016/j.addbeh.2014.04.017] [Medline: 24837754]
48. Hymowitz N, Cummings KM, Hyland A, Lynn WR, Pechacek TF, Hartwell TD. Predictors of smoking cessation in a cohort of adult smokers followed for five years. Tob Control 1997;6(Suppl 2):S57-S62 [FREE Full text] [doi: 10.1136/tc.6.suppl_2.s57] [Medline: 9583654]
49. Stolz D, Scherr A, Seiffert B, Kuster M, Meyer A, Fagerström KO, et al. Predictors of success for smoking cessation at the workplace: a longitudinal study. Respiration 2014;87(1):18-25 [FREE Full text] [doi: 10.1159/000346646] [Medline: 23594795]
50. Kale D, Gilbert HM, Sutton S. Are predictors of making a quit attempt the same as predictors of 3-month abstinence from smoking? Findings from a sample of smokers recruited for a study of computer-tailored smoking cessation advice in primary care. Addiction 2015 Oct;110(10):1653-1664. [doi: 10.1111/add.12972] [Medline: 25939254]
51. Reid JL, Hammond D, Boudreau C, Fong GT, Siahpush M, ITC Collaboration. Socioeconomic disparities in quit intentions, quit attempts, and smoking abstinence among smokers in four western countries: findings from the International Tobacco Control Four Country Survey. Nicotine Tob Res 2010 Oct;12(Suppl 1):S20-S33 [FREE Full text] [doi: 10.1093/ntr/ntq051] [Medline: 20889477]
52. Ubhi HK, Kotz D, Michie S, van Schayck OCP, West R. A comparison of the characteristics of iOS and Android users of a smoking cessation app. Transl Behav Med 2017 Jun;7(2):166-171 [FREE Full text] [doi: 10.1007/s13142-016-0455-z] [Medline: 28168609]
53. Benowitz NL, Bernert JT, Foulds J, Hecht SS, Jacob P, Jarvis MJ, et al. Biochemical verification of tobacco use and abstinence: 2019 update. Nicotine Tob Res 2020 Jun 12;22(7):1086-1097. [doi: 10.1093/ntr/ntz132] [Medline: 31570931]
54. Connor Gorber S, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. Nicotine Tob Res 2009 Jan;11(1):12-24. [doi: 10.1093/ntr/ntn010] [Medline: 19246437]

Abbreviations

CASC: CureApp Smoking Cessation
CO: carbon monoxide
CPD: cigarettes per day
FDA: Food and Drug Administration
IRB: Institutional Review Board
ITT: intention to treat
NRT: nicotine replacement therapy
OR: odds ratio
PPA: point prevalence abstinence
RCT: randomized controlled trial
RR: relative risk
USCPG: United States Clinical Practice Guideline

Edited by G Eysenbach; submitted 06.11.20; peer-reviewed by K Masaki, J Heffner, J Steinerman, M Bardus; comments to author 27.11.20; revised version received 11.01.21; accepted 21.01.21; published 15.02.21.

Please cite as:
Marler JD, Fujii CA, Galanko JA, Balbierz DJ, Utley DS
Durability of Abstinence After Completing a Comprehensive Digital Smoking Cessation Program Incorporating a Mobile App, Breath Sensor, and Coaching: Cohort Study
J Med Internet Res 2021;23(2):e25578
URL: http://www.jmir.org/2021/2/e25578/
doi:10.2196/25578
PMID:33482628

Original Paper

Access to Technology and Preferences for an mHealth Intervention to Promote Medication Adherence in Pediatric Acute Lymphoblastic Leukemia: Approach Leveraging Behavior Change Techniques

Mallorie B Heneghan[1], MD, MSc; Tasmeen Hussain[2], MD, MPH; Leonardo Barrera[3], MPH; Stephanie W Cai[4], MD; Maureen Haugen[5], MSc; Elaine Morgan[5,6], MD; Jenna Rossoff[5,6], MD; Joanna Weinstein[5,6], MD; Nobuko Hijiya[7], MD; David Cella[8], PhD; Sherif M Badawy[5,6], MD, MSc

[1]Division of Pediatric Hematology/Oncology, Department of Pediatrics, University of Utah, Salt Lake City, UT, United States

[2]Division of Internal Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[3]Mary Ann & J. Milburn Smith Child Health Research, Outreach, and Advocacy Center, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, United States

[4]Department of Obstetrics & Gynecology, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[5]Division of Hematology, Oncology and Stem Cell Transplantation, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, United States

[6]Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[7]Department of Pediatrics, Columbia University Medical Center, New York, NY, United States

[8]Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

Corresponding Author:
Mallorie B Heneghan, MD, MSc
Division of Pediatric Hematology/Oncology
Department of Pediatrics
University of Utah
100 N Mario Capecchi Drive
Salt Lake City, UT, 84113
United States
Phone: 1 801 662 4700
Email: Mallorie.Heneghan@hsc.utah.edu

Abstract

Background: Suboptimal adherence to 6-mercaptopurine (6-MP) is prevalent in pediatric acute lymphoblastic leukemia (ALL) and associated with increased risk of relapse. Rapid uptake of personal technology makes mobile health (mHealth) an attractive platform to promote adherence.

Objective: Study objectives were to examine access to mobile technology and preferences for an mHealth intervention to improve medication adherence in pediatric ALL.

Methods: A cross-sectional survey was administered in oncology clinic to parents of children with ALL as well as adolescents and young adults (AYAs) with ALL receiving maintenance chemotherapy.

Results: A total of 49 parents (median age [IQR] 39 [33-42] years; female 76% [37/49]) and 15 patients (median age [IQR] 17 [16-19]; male 80% [12/15]) participated. All parents and AYAs owned electronic tablets, smartphones, or both. Parents' most endorsed mHealth app features included a list of medications (71%, 35/49), information about 6-MP (71%, 35/49), refill reminders (71%, 35/49), and reminders to take 6-MP (71%, 35/49). AYAs' most endorsed features included refill reminders (73%, 11/15), reminders to take 6-MP (73%, 11/15), and tracking 6-MP (73%, 11/15).

Conclusions: Parents and AYAs reported ubiquitous access to mobile technology and strong interest in multiple adherence-specific mHealth app features. Parents and AYAs provided valuable insight into preferred features for a multifunctional behavioral intervention (mHealth app) to promote medication adherence in pediatric ALL.

(J Med Internet Res 2021;23(2):e24893) doi:10.2196/24893

KEYWORDS

acute lymphoblastic leukemia; medication adherence; behavior change technique; oral chemotherapy; mHealth; patient-centered

Introduction

Acute lymphoblastic leukemia (ALL) is the most common pediatric malignancy [1]. While 50 years ago ALL was invariably fatal, with modern chemotherapy plans, the vast majority of children will achieve remission within 2 months of starting therapy [2,3]. However, further treatment is needed to prevent relapse [2,3]. Despite significant improvements in therapy, 15%-20% of patients still relapse, and cure rates after relapse are considerably less favorable between 20% and 50% [2]. Adding 18-32 months of a low-intensity "maintenance" phase with multiple oral chemotherapy medications, including daily oral 6-mercaptopurine (6-MP), has decreased relapse rates [4].

Low adherence to oral 6-MP has been identified as a significant challenge [5-7], because even minimally suboptimal adherence has been associated with an increased risk of relapse [8-10]. A clinical trial from the Children's Oncology Group using electronic monitoring devices found that 44% of patients had adherence rates under 95%, which was associated with a 2.7-fold increased risk of ALL relapse compared to those with adherence of 95% or higher [9]. Efforts to promote medication adherence are essential to preventing relapse and improving survival.

Medication adherence is a complex behavior [11]. A variety of interventions have been proposed to improve medication adherence, but behavioral interventions are often complicated and multifaceted with no clear underlying theoretical framework [12]. When these interventions reach their intended goal, disentangling the critical components can be difficult, limiting dissemination [13]. The behavior change wheel (BCW) provides a framework to guide effective development and implementation of behavior change interventions. At the center of the BCW health behaviors are conceptualized in terms of capability, opportunity, and motivation (the COM-B framework) [14,15]. The second layer of the BCW allows researchers to link a conceptualized behavior to 9 intervention functions (education, persuasion, incentivization, coercion, training, enablement, modeling, environmental restructuring, and restrictions). The 9 intervention functions have been mapped to specific behavior change techniques (BCTs) within the Behavior Change Technique Taxonomy (BCTT) [15-17]. The BCTT is a classification system for BCTs developed using a series of expert consensus exercises aimed at improving the reliability and specificity of behavior change interventions by allowing for identification of effective components of behavioral interventions [16]. The BCTT can help researchers identify the active components of interventions that should be implemented to promote a specific element of a behavior. Medication adherence behavior can be better understood and optimized using BCTT.

Mobile health (mHealth) is the use of mobile and wireless applications (eg, SMS text messaging, apps, wearable devices, remote sensing, and the use of social media) in the delivery of health-related services [18,19]. mHealth is an attractive platform for implementing BCTTs because access to mobile technology is widespread. Furthermore, leveraging mHealth interventions is key to optimizing health outcomes in patients with chronic medical conditions during the current COVID-19 pandemic, including ALL [20]. The most recent Pew Research Institute survey reported that 96% of adult US residents own a mobile phone and 81% own a smartphone [21]. Similarly, 95% of teens (age 13-17) reported owning or having access to a smartphone and almost half reported "being online constantly" [22]. The Pew Research Institute survey was designed to capture a sample of households whose demographics mirror the United States nationally but does not necessarily capture individuals with chronic medical conditions. Specifically, families caring for a child with cancer face multiple challenges including material hardship [23] which may affect their access to technology. Ensuring pediatric oncology parents and adolescents and young adults (AYAs) with cancer have access to technology is a prerequisite when considering mHealth interventions.

Access, connectivity, and engagement are all prerequisites for a successful mHealth intervention. Several mHealth apps to promote adherence are available commercially [24] and a pilot study of a medication reminder app demonstrated the feasibility and perceived usefulness of using mHealth in AYAs with cancer [25]. However, mHealth apps have not been yet adopted by pediatric oncology parents [26]. While technology may provide a novel means of implementing behavioral intervention, previous research reports low mHealth use in parents of children with cancer, with Google and WebMD being cited as the most helpful mHealth apps/websites to care for their child [26]. Additionally, while adolescents in general are avid smartphone users, only 2% of teens report frequently using an mHealth app [27]. Being connected to and comfortable with mobile technology are necessary components but are not sufficient for the success of mHealth apps. Successful mHealth apps also need to engage users [28].

Involving users in the development process early promotes engagement [28,29]. User-centered app design is a method for designing mobile apps which starts with a needs assessment followed by iterative cycles involving the intended end user [30]. mHealth apps developed with end user input from the early conceptualization of the intervention through development, deployment, testing, implementation and dissemination are more likely to be perceived by users as useful and easy to use [30]. Engagement in mHealth interventions requires both access to technology and interest in its use [28]. Therefore, understanding patients' and parents' access and interest in technology-based (mHealth) interventions is the first step in developing an effective tool to promote medication adherence during maintenance therapy and optimize health outcomes in pediatric ALL.

In this study, we aimed to conduct an mHealth adherence app needs assessment as the first step in user-centered design process. Our first objective was to evaluate access to and comfort with mobile technology among parents who have a

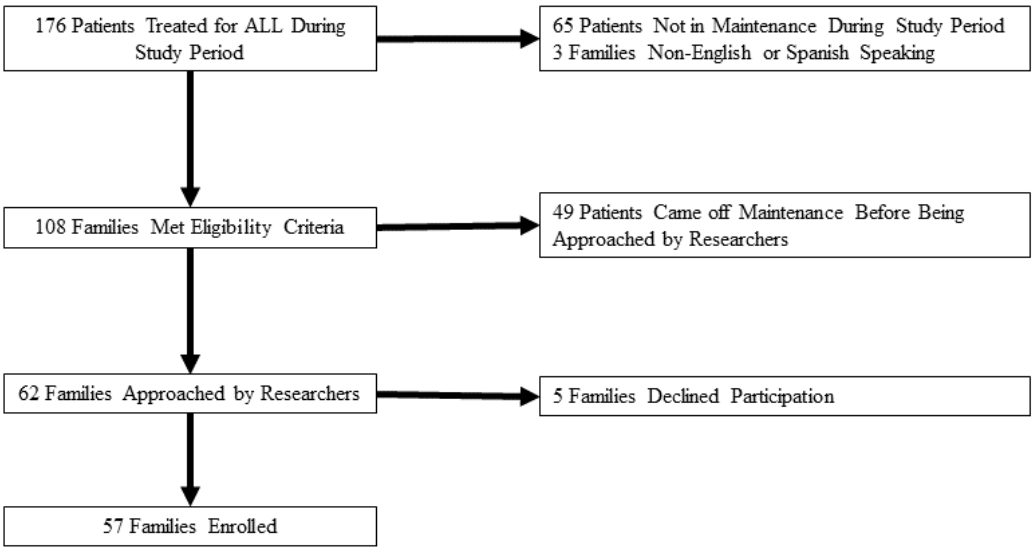
child with ALL and AYAs with ALL. Second, we aimed to evaluate interest and preferences for a technology-based (mHealth) intervention (ie, mobile app) to promote adherence to oral chemotherapy during maintenance therapy. We hypothesized that both AYAs with ALL and parents of children with ALL could have variable access to mobile technology and high interest in a mHealth app to promote adherence. We also hypothesized that AYAs with ALL and parents of children with ALL will have different preferences and priorities for mHealth.

Methods

Participants and Survey Administration

We completed a cross-sectional study at a single institution. Eligibility criteria included English or Spanish speaking parents of children with ALL (ages 1-18 years) and patients (ages 12-24 years) with ALL in remission and actively receiving 6-MP as part of the maintenance phase of therapy. Potential study participants were approached before or after regularly scheduled outpatient oncology clinic appointments from November 2017 to March 2019 (Figure 1). Surveys were administered through REDCap using study electronic tablets. Written consent was obtained from all parent participants and patient participants aged over 18. Parent consent and patient assent were obtained for patient participants aged 12-17. The Institutional Review Board at the Ann and Robert H. Lurie Children's Hospital of Chicago approved the study and all procedures were conducted in accordance with the current version of the Helsinki Declaration.

Figure 1. Enrollment flowchart.



Data Collection

Data were collected on electronic tablets through REDCap supported by the Northwestern University Clinical and Translational Sciences (NUCATS) Institute. Medical chart review of enrolled patients was conducted to collect information on ALL disease characteristics and patient characteristics.

Study Measures

Our study instrument included 65 items assessing technology access, mHealth preferences, and demographics. All portions were developed based on available evidence in the literature related to technology-based interventions and medication adherence for chronic conditions in both adult and pediatric populations that our group has previously published [31-34].

The technology access portion of the survey included 7 questions about access to electronic devices, as well as questions evaluating monthly data plans, SMS text message and call limits, quality and speed of broadband 4G, or home internet signal strength. An additional 10 questions asked about technology use habits and 12 questions assessed participants' comfort with technology.

The mHealth portion of the survey included 8 yes/no questions, 1 rank order question, and 6 multiple-choice questions that evaluated interest in general mHealth app features and notification preferences, all of which have been reported by our group previously [33]. Our multidisciplinary team of pediatric hematologists, behavioral scientists, and health educators designed these questions, which were later pretested to ensure they were age appropriate for adolescents. Participants were asked to rank the most important features from 1 to 4, with 1 being the most important and 4 being the least important feature. We also developed a list of 20 mHealth features aimed at promoting medication adherence, informed by the Disease

Management and Barriers Interview [35], which was developed to identify barriers to adherence, and BCTs [16]. mHealth features were mapped to the BCTT by 2 researchers (MBH and SWC) independently [16,17]. Discrepancies were resolved via discussion with a third researcher (SB), which lead to 100% consensus agreement. According to the BCTT, interventions can have more than 1 function and when agreed upon by 3 researchers (MBH, SWC, and SMB), features were mapped to more than 1 technique. We included the final agreed upon classification of questions in the "Results" section.

Statistical Analysis

Descriptive statistics for categorical data were reported in frequencies and percentages. Nonparametric stringent statistics were used, because study sample data were not normally distributed. Wilcoxon rank-sum test, Kruskal–Wallis test, chi-square, and Fisher exact tests were performed when appropriate to determine significant association (P<.05) among variables. Spearman Rho correlations (r_s) were calculated to examine the relationship between relevant continuous variables. Statistical analyses were performed using STATA 15.1 (Stata Statistical Software, Release 15.1; StataCorp LP) and MS Excel for Mac 2019 (version 16.27).

Results

Participant Characteristics

Of the 62 eligible families approached, 57 (92%) consented to participation (49 parents and 15 patients, including 8 dyads) and completed the study survey (Figure 1). The median age of the participating parents was 39 years (IQR 33-42 years) and 76% were female (37/49). Parents reported on children with a median age of 6 years (IQR 5-10 years). Patient participants had a median age of 17 years (IQR 16-19 years) and 80% (12/15) were male. All participants' characteristics are summarized in Table 1.

Table 1. Participants' characteristics.

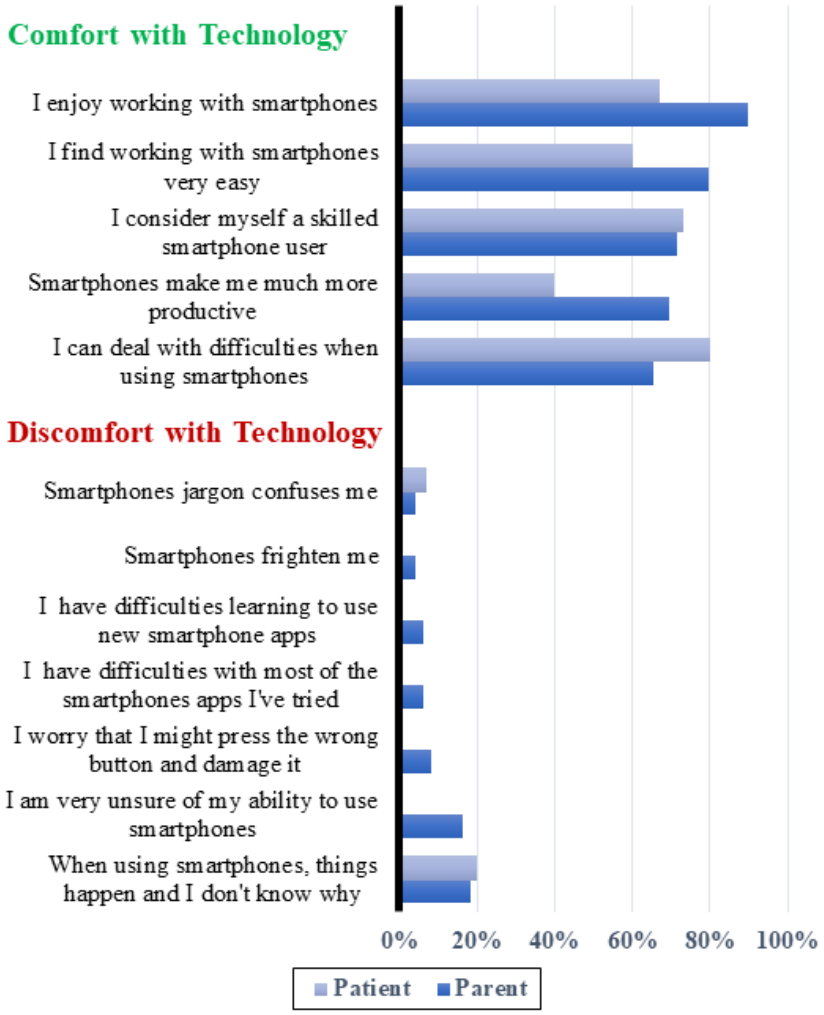
Characteristics	Parent (N=49)	Patient (N=15)
Age, median (IQR)	39 (33-42)	17 (16-19)
Sex, n (%)		
Female	37 (76)	3 (20)
Male	12 (24)	12 (80)
Race, n (%)		
White	25 (51)	5 (33)
Black/African American	2 (4)	1 (7)
Hispanic or Latino	18 (37)	9 (60)
Asian	4 (8)	2 (13)
Highest level of education, n (%)		
Less than high school	8 (16)	8 (53)
High-school diploma	6 (12)	5 (33)
Some college, but no degree	10 (20)	2 (13)
Associate degree	6 (12)	—
Bachelor's degree or higher	19 (39)	—

Access to and Comfort With Mobile Technology

All parents and patients owned an electronic tablet, a smartphone, or both. Most parents and patients (96% [47/49] and 100% [15/15]) owned smartphones, mainly iPhones (both 73% [36/49, 11/15]), as well as tablets (84% [41/49] and 67% [10/15]) and laptops (63% [31/49] and 67% [10/15]), respectively. Most parents and patients had unlimited plans for SMS text messaging (98% [48/49] and 93% [14/15]) and data (69% [34/49] and 67% [10/15]), as well as a fast home internet connection (92% [45/49] and 93% [14/15]), respectively.

Most parents and patients agreed with statements about being comfortable with technology (Figure 2). Only a few parents and patients agreed with statements expressing discomfort with technology. In particular, 18% [9/49] of parents and 20% [3/15] of patients agreed that "When using smartphones things happen, and I don't know why." Technology comfort questionnaire data are summarized in Multimedia Appendix 1.

Figure 2. Parents' and patients' comfort and discomfort with technology.



Interest in mHealth Features

All 8 proposed mHealth features were endorsed by over 50% of parents (25/49) and patients [8/15] (Table 2). The median number of features endorsed by both parents and patients was 7 (IQR 6-8). For parents, the most endorsed features were having the ability to review laboratory results (98%, 48/49), information about ALL medications (96%, 47/49), and information about ALL (92%, 45/49). By contrast, the features most endorsed by patients were daily 6-MP reminders (93%, 14/15), 6-MP tracking (93%, 14/15), and education about ALL (93%, 14/15). Patients were significantly more likely to express interest in 6-MP reminders (93% [14/15] versus 88% [43/49], *P*=.02) and 6-MP tracking (93% [14/15] versus 86% [42/49], *P*=.01), when compared to parents, respectively.

Table 2. Frequency of participants' reported interest in general mHealth features.

Question	Parents (N=49), n (%)	Patients (N=15), n (%)	*P* value
Remind you to take/give your child 6-MP[a] every day?	43 (88)	14 (93)	.02[c]
Record when you/your child takes 6-MP every day?	42 (86)	14 (93)	.01[c]
Provide encouraging messages when you/your child takes 6-MP?	32 (65)	8 (53)	.87
Send a text message reminder when you/your child hasn't taken 6-MP?	40 (82)	12 (80)	.69
Virtually connect you to other patients with ALL[b] and their families?	31 (63)	9 (60)	.16
Provide information about ALL?	45 (92)	11 (73)	.12
Provide information about ALL medications (such as 6-MP and steroid medications) and how they work?	47 (96)	14 (93)	.37
Show the results of your/your child's blood tests?	48 (98)	13 (87)	.49

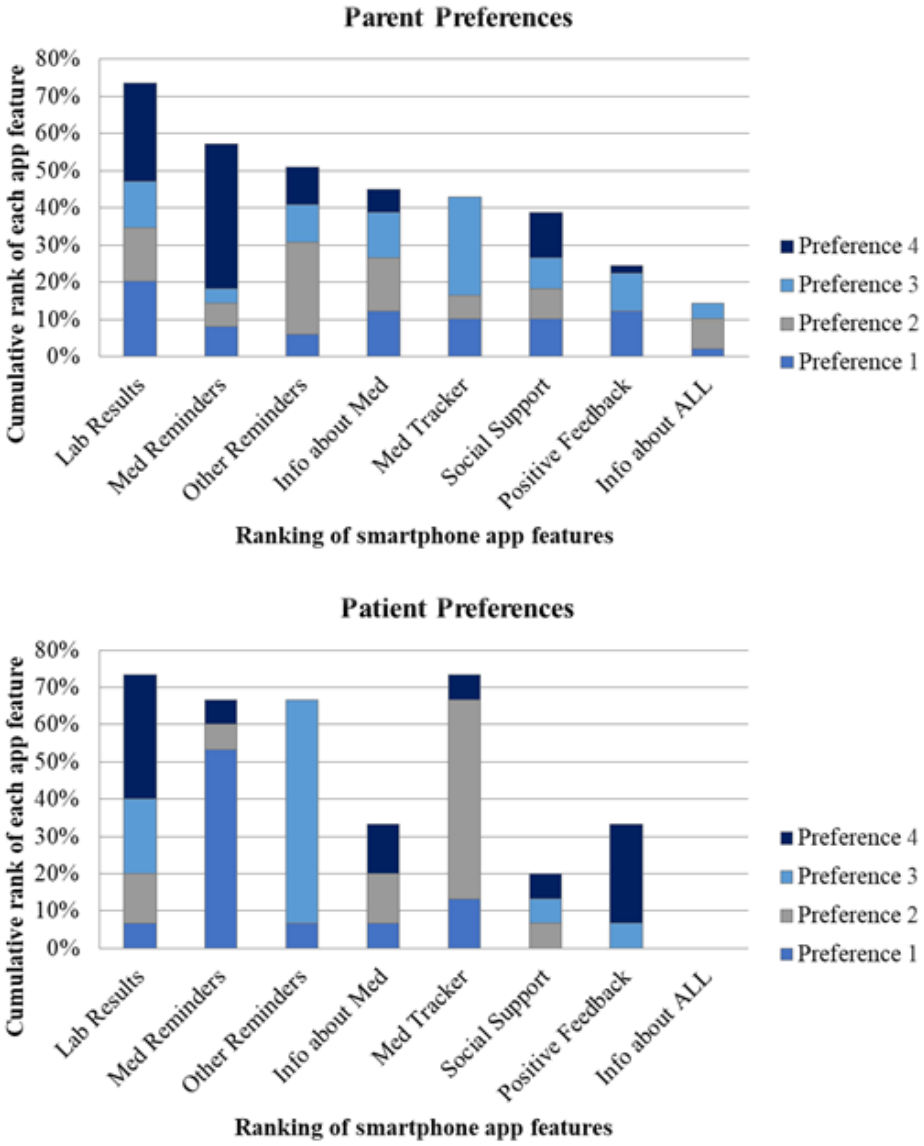
[a]6-MP: 6 mercaptopurine.

[b]ALL: acute lymphoblastic leukemia.

[c]*P* value <.05 was statistically significant.

The cumulative ranking of the proposed ALL smartphone app features among parents and AYAs are summarized and illustrated in Figure 3. The ability to review laboratory results (20% [10/49]) was ranked most important most frequently among parents, followed by education about ALL medications (12% [6/49]) and positive feedback (12% [6/49]). By contrast, patients prioritized medication reminders (53% [8/15]) and medication tracker (13% [2/15]).

Figure 3. Participants' cumulative ranking of smartphone adherence app features.



mHealth Features to Support Medication Adherence

When asked specifically about their interest in mHealth app features to support medication adherence, parents and patients endorsed a median of 13 features (IQR 8-18) and 11 features (IQR 4-18), respectively (Table 3; *P*=.72). Parents and patients endorsed adherence-promoting app features, mapped to BCTT, which are summarized in Table 3. Parents' top endorsed adherence-promoting mHealth app features included information about what 6-MP does (71%, 35/49), reminders to refill 6-MP (71%, 35/49), reminders to take 6-MP (71%, 35/49), and a list of medications (71%, 35/49). Patients most frequently expressed interest in reminders to refill 6-MP (73%, 11/15), reminders to take 6-MP (73%, 11/15), and a way to track 6-MP (73%, 11/15). Both parents and patients were most interested in mHealth app features that mapped to prompts and cues (BCTT 7.1). Patients preferred features that utilized self-monitoring behavior (BCTT 2.3), while parents were more interested in instructions on how to perform a behavior (BCTT 4.1) and information about health consequences (BCTT 5.1).

Table 3. Participants' interest in mHealth behavior change techniques focused on promoting medication adherence.

App features by behavior change technique taxonomy (number and label)	Parent (N=49), n (%)	Patient (N=15), n (%)
1.2. Problem solving		
Help getting to ALL[a] appointments	19 (39)	3 (20)
Help paying for ALL appointments	24 (49)	7 (47)
2.3. Self-monitoring behavior		
A way to track 6-MP[b] administration	31 (63)	11 (73)
2.6. Biofeedback		
Easier access to my child's health record	29 (59)	9 (60)
3.1. Social support		
Virtually connect to other patients with ALL and their families	23 (47)	6 (40)
Information for friends and family about my child's ALL and 6-MP	26 (53)	8 (53)
4.1. Instructions on how to perform a behavior		
A list of medications on the app	35 (71)	10 (67)
Information about how to take 6-MP	24 (49)	7 (47)
Reminders and summaries of doctor's instructions	34 (69)	9 (60)
5.1. Information about health consequences		
Information about what 6-MP does on an app	35 (71)	10 (67)
List of side effects of 6-MP on an app	33 (67)	10 (67)
Information about why it is important to take 6-MP	31 (63)	9 (60)
7.1. Prompts/cues		
App reminders to refill my child's 6-MP	35 (71)	11 (73)
App reminders to give 6-MP to my child	35 (71)	11 (73)
Reminders to take 6-MP when at school or away from home	31 (63)	10 (67)
Reminders and summaries of doctor's instructions	34 (69)	9 (60)
9.1. Credible source		
Provide information about ALL?	33 (67)	8 (53)
Information about what my 6-MP does on an app	35 (71)	10 (67)
Information for friends and family about my child's ALL and 6-MP	26 (53)	8 (53)
12.1. Restructuring the physical environment		
An easier way to contact my child's ALL doctor	25 (51)	7 (47)
Easier access to my child's health record	29 (59)	9 (60)
12.2. Restructuring the social environment		
An easier way to contact my child's ALL doctor	25 (51)	7 (47)
A way for me to meet other patients with ALL	23 (47)	5 (33)
A way to contact community centers with ALL resources	25 (51)	7 (47)

[a]ALL: acute lymphoblastic leukemia.

[b]6-MP: 6 mercaptopurine.

Discussion

Our study contributes to the existing literature on medication adherence in pediatric ALL by reporting on the potential role of mHealth interventions to optimize adherence behavior in this population. Using a cross-sectional survey design, parents and patients with ALL self-reported on their access to technology as well as preferences for mHealth intervention features. We found that all participants owned either a smartphone or an electronic tablet with the majority owning both. Most parents and patients were comfortable using mobile technology. All participants were interested in at least one mHealth feature and all 8 proposed features were endorsed by more than half of both the parent and patient participants. When asked to prioritize

these features, parents most commonly rated access to laboratory results, while patients prioritized medication reminders and a medication tracker. Additionally, all parents and patients expressed strong interest in features aimed at promoting medication adherence.

The high rates of technology access are comparable to what has been reported in the general population for AYAs [21,22]. Previous work by Mueller et al [26] demonstrated similar rates of technology access among parents of oncology patients and interest in mHealth [26]. Our study adds to existing literature by demonstrating ubiquitous technology access and interest in mHealth among both parents of children with leukemia and AYAs with leukemia. A unique feature in our study is the increased racial/ethnic diversity in our sample as well as higher completion rate, both of which increase the generalizability of our findings compared to what has been previously reported.

Our findings suggest previously reported low rates of mHealth use by adolescents [27] and parents of children with cancer [26] are not due to a lack of interest in mHealth and represents the first step in user-centered design. Numerous smartphone mHealth apps from commercial vendors are available for use [24,36]. A recent international survey of key stakeholders (ie, clinician, patient organizations, and experts) reported that current mHealth apps seldom meet patients' expectations and needs because they were not developed with patients in mind [37]. This may account for why most downloaded mobile apps are retained less than a day [38]. Early end user involvement in mobile app development has been associated with increased engagement [28,29]. However, review of the mHealth apps available for download on the iTunes and Google Play app stores confirmed that only a handful of mHealth apps were designed for patients, parents, or both [24]. Further, the majority of these mHealth apps were solely designed for educational purposes, not disease management or medication adherence [24]. Assessing parents' and patients' interest in mHealth is the beginning of the continuum of user-centered design. Future efforts should focus on ongoing end user feedback on prototypes and also include features aimed at promoting engagement such as customization, avatars, incentives, and gamification [39].

Previous publications of smartphone mHealth app preferences in pediatric oncology did not focus specifically on medication adherence promotion, which we addressed in this study. Our results are consistent with what has been reported in other complex chronic conditions including cystic fibrosis [40], diabetes [41,42], and sickle cell disease [33], suggesting that a multicomponent smartphone mHealth app may represent a novel intervention to promote oral chemotherapy adherence in patients with ALL. Furthermore, previous systematic reviews demonstrated that interventions with more incorporated BCTs had a larger effect size than other interventions incorporating fewer BCTs [43]. These findings support our patients' and parents' interest in multiple key BCTs, which is a unique feature of our study.

AYAs and parents expressed unique priorities. AYAs have different mHealth needs, given their developmental stages and generational differences from their parents. The unique preferences of AYAs are consistent with a review of mHealth interventions suggesting that AYAs may have higher tolerance for SMS text message fatigue than adults, and advocating for adolescent involvement in the development process [32,44]. Parents' and adolescents' preferences for an mHealth intervention to promote adherence have not been directly compared previously. However, several studies have found that barriers to adherence vary throughout the lifespan and are unique among adolescents [45,46]. The expressed preference for a medication tracker is consistent with previous studies that suggest organization is important for adolescents. Adolescents expressed low interest in social networking features which is inconsistent with theories on development that suggest adolescents are particularly attuned to peer input and engage in social networks at high rates [46]. In addition, in a qualitative study of adolescents with asthma participants expressed that having the ability to interact with peers on an adherence app would promote adherence. We hypothesize that the lower prevalence of ALL and potential stigma associated with being a patient with pediatric cancer [47] may impact interest in social networking. These differences could be informative and important to consider during a user-centered mHealth app development process to improve engagement with both patients and parents. In other words, having core features for an mHealth ALL app with slightly modified patient and parent versions may be beneficial. Understanding patients' and parents' preferences through continued user-centered design approach is vital for the success of mHealth-delivered BCTs [28].

Our study has several strengths. First, we had a high enrollment and completion rate of our survey in a racially/ethnically diverse sample, while previous work evaluating access to technology was limited due to a low response rate. Second, we provide the most comprehensive evaluation of access to technology, evaluating multiple modes of technology as well as comfort with and barriers to technology, including mobile technology, SMS text message, and data plans. Third, we are the first to report on and compare patients' and parents' preferences for an ALL mHealth app as a behavioral intervention to promote medication adherence. Finally, we used an established theoretical framework for behavioral intervention development, the BCW/COM-B, and classified our proposed interventions using the BCTT. Using the BCW/COM-B and BCTT allows us to create a targeted intervention while building on previous work in behavior change.

Our study has some limitations worth mentioning. Despite the high enrollment rate and diverse patient population, the generalizability of this study is limited due to data being collected from a single institution with a relatively small sample size. While we adapted validated survey items when possible, not all of our survey items were validated, yet they have been used in other published studies [33,34]. Because we adapted existing survey items, we were not able to provide a more exhaustive list of potential interventions, and we acknowledge that other potential interventions were not included. Finally, we did not ask about current or past use of mHealth apps which could impact participants' perceptions of mHealth and their preferences.

In conclusion, parents and patients reported ubiquitous access to mobile technology and high levels of interest in general ALL

mHealth app features and features intended to promote medication adherence. Our findings highlight the need for the development of a user-centered mHealth intervention (mobile app) to promote medication adherence among children and AYAs with ALL and their parents. Designing 2 separate app versions for each group of end users, parents and patients, may be needed to optimize engagement with the app as an intervention and improve care delivery, and ultimately health outcomes in this vulnerable population of pediatric ALL.

Data Availability Statement

The data sets generated during or analyzed during this study are available from the corresponding author on reasonable request.

Acknowledgments

This project was supported by grants from the National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH) (K23HL150232) (PI: SB) and the Agency for Healthcare Research and Quality (AHRQ) (K12HS023011) (PI: SB). The project was also supported by the Northwestern University Clinical and Translational Sciences Institute (NUCATS) (PI: SB) through the National Institutes of Health's National Center for Advancing Translational Sciences, UL1TR001422. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute (NHLBI), the National Institutes of Health (NIH), or the Agency for Healthcare Research and Quality (AHRQ).

Authors' Contributions

TH and SMB designed the research study; MBH, TH, LB, SWC, and SMB performed the research and collected the data; MBH analyzed the data; MBH and SMB interpreted the data and drafted the paper; TH, LB, SWC, MH, EM, JR, JW, NH, and DC critically revised the paper, and all authors approved the submitted final version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participants' comfort and discomfort with mobile technology.
[DOCX File , 14 KB - jmir_v23i2e24893_app1.docx]

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA A Cancer J Clin 2019 Jan 08;69(1):7-34. [doi: 10.3322/caac.21551]
2. Hunger SP, Mullighan CG. Acute Lymphoblastic Leukemia in Children. N Engl J Med 2015 Oct 15;373(16):1541-1552. [doi: 10.1056/NEJMra1400972] [Medline: 26465987]
3. Hunger SP, Loh ML, Whitlock JA, Winick NJ, Carroll WL, Devidas M, COG Acute Lymphoblastic Leukemia Committee. Children's Oncology Group's 2013 blueprint for research: acute lymphoblastic leukemia. Pediatr Blood Cancer 2013 Jun 19;60(6):957-963 [FREE Full text] [doi: 10.1002/pbc.24420] [Medline: 23255467]
4. Childhood ALL Collaborative Group. Duration and intensity of maintenance chemotherapy in acute lymphoblastic leukaemia: overview of 42 trials involving 12 000 randomised children. Lancet 1996 Jun 29;347(9018):1783-1788. [doi: 10.1016/s0140-6736(96)91615-3] [Medline: 8667921]
5. Heneghan M, Hussain T, Barrera L, Cai SW, Haugen M, Duff A, et al. Applying the COM-B model to patient-reported barriers to medication adherence in pediatric acute lymphoblastic leukemia. Pediatr Blood Cancer 2020 May;67(5):e28216. [doi: 10.1002/pbc.28216] [Medline: 32068338]
6. Gupta S, Bhatia S. Optimizing medication adherence in children with cancer. Current Opinion in Pediatrics 2017;29(1):41-45. [doi: 10.1097/mop.0000000000000434]
7. Landier W, Chen Y, Hageman L, Kim H, Bostrom BC, Casillas JN, et al. Factors associated with nonadherence to oral 6-mercaptopurine (6MP) in children with acute lymphoblastic leukemia (ALL): A report from Children's Oncology Group (COG) study AALL03N1. J Clin Oncol 2014 May 20;32(15_suppl):10013. [doi: 10.1200/jco.2014.32.15_suppl.10013]
8. Landier W, Hageman L, Chen Y, Kornegay N, Evans WE, Bostrom BC, et al. Mercaptopurine Ingestion Habits, Red Cell Thioguanine Nucleotide Levels, and Relapse Risk in Children With Acute Lymphoblastic Leukemia: A Report From the Children's Oncology Group Study AALL03N1. J Clin Oncol 2017 May 20;35(15):1730-1736 [FREE Full text] [doi: 10.1200/JCO.2016.71.7579] [Medline: 28339328]
9. Bhatia S, Landier W, Hageman L, Chen Y, Kim H, Sun C, et al. Systemic Exposure to Thiopurines and Risk of Relapse in Children With Acute Lymphoblastic Leukemia: A Children's Oncology Group Study. JAMA Oncol 2015 Jun 01;1(3):287-295 [FREE Full text] [doi: 10.1001/jamaoncol.2015.0245] [Medline: 26181173]

10. Bhatia S, Landier W, Hageman L, Kim H, Chen Y, Crews KR, et al. 6MP adherence in a multiracial cohort of children with acute lymphoblastic leukemia: a Children's Oncology Group study. Blood 2014 Oct 09;124(15):2345-2353 [FREE Full text] [doi: 10.1182/blood-2014-01-552166] [Medline: 24829202]
11. Peipert JD, Badawy SM, Baik SH, Oswald LB, Efficace F, Garcia SF, et al. Development of the NIH Patient-Reported Outcomes Measurement Information System (PROMIS) Medication Adherence Scale (PMAS). PPA 2020 Jun;Volume 14:971-983. [doi: 10.2147/ppa.s249079]
12. Badawy SM, Shah R, Beg U, Heneghan MB. Habit Strength, Medication Adherence, and Habit-Based Mobile Health Interventions Across Chronic Medical Conditions: Systematic Review. J Med Internet Res 2020 Apr 28;22(4):e17883 [FREE Full text] [doi: 10.2196/17883] [Medline: 32343250]
13. Michie S, West R, Sheals K, Godinho CA. Evaluating the effectiveness of behavior change techniques in health-related behavior: a scoping review of methods used. Transl Behav Med 2018 Mar 01;8(2):212-224 [FREE Full text] [doi: 10.1093/tbm/ibx019] [Medline: 29381786]
14. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
15. Michie SAL, West R. The Behaviour Change Wheel: A Guide to Developing Interventions. London, UK: Silverback Publishing; 2014.
16. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug 20;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
17. Michie S, Carey RN, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. From Theory-Inspired to Theory-Based Interventions: A Protocol for Developing and Testing a Methodology for Linking Behaviour Change Techniques to Theoretical Mechanisms of Action. Ann Behav Med 2018 May 18;52(6):501-512 [FREE Full text] [doi: 10.1007/s12160-016-9816-6] [Medline: 27401001]
18. Borrelli B, Ritterband LM. Special issue on eHealth and mHealth: Challenges and future directions for assessment, treatment, and dissemination. Health Psychol 2015 Dec;34S:1205-1208. [doi: 10.1037/hea0000323] [Medline: 26651461]
19. Radovic A, Badawy SM. Technology Use for Adolescent Health and Wellness. Pediatrics 2020 May;145(Suppl 2):S186-S194. [doi: 10.1542/peds.2019-2056G] [Medline: 32358210]
20. Badawy S, Radovic A. Digital Approaches to Remote Pediatric Health Care Delivery During the COVID-19 Pandemic: Existing Evidence and a Call for Further Research. JMIR Pediatr Parent 2020 Jun 25;3(1):e20049 [FREE Full text] [doi: 10.2196/20049] [Medline: 32540841]
21. Pew Research Center. Mobile Fact Sheet. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-01-26]
22. Lenhart A. Teen, Social Media and Technology Overview. 2015. URL: https://www.slideshare.net/eraser/teens-social-media-technology-overview-2015 [accessed 2021-01-26]
23. Bilodeau M, Ma C, Al-Sayegh H, Wolfe J, Bona K. Household material hardship in families of children post-chemotherapy. Pediatr Blood Cancer 2018 Jan 20;65(1):e26743 [FREE Full text] [doi: 10.1002/pbc.26743] [Medline: 28941160]
24. Jupp JCY, Sultani H, Cooper CA, Peterson KA, Truong TH. Evaluation of mobile phone applications to support medication adherence and symptom management in oncology patients. Pediatr Blood Cancer 2018 Nov 26;65(11):e27278. [doi: 10.1002/pbc.27278] [Medline: 29943893]
25. Wu Y, Linder L, Kanokvimankul P, Fowler B, Parsons B, Macpherson C, et al. Use of a Smartphone Application for Prompting Oral Medication Adherence Among Adolescents and Young Adults With Cancer . ONF 2018 Jan 1;45(1):69-76. [doi: 10.1188/18.onf.69-76]
26. Mueller EL, Cochrane AR, Bennett WE, Carroll AE. A survey of mobile technology usage and desires by caregivers of children with cancer. Pediatr Blood Cancer 2018 Nov 17;65(11):e27359. [doi: 10.1002/pbc.27359] [Medline: 30015371]
27. Wartella E, Rideout V, Montague H, Beaudoin-Ryan L, Lauricella A. Teens, Health and Technology: A National Survey. MaC 2016 Jun 16;4(3):13-23. [doi: 10.17645/mac.v4i3.515]
28. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
29. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
30. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
31. Badawy SM, Kuhns LM. Texting and Mobile Phone App Interventions for Improving Adherence to Preventive Behavior in Adolescents: A Systematic Review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]

32. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text Messaging and Mobile Phone Apps as Interventions to Improve Adherence in Adolescents With Chronic Health Conditions: A Systematic Review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
33. Badawy SM, Thompson AA, Liem RI. Technology Access and Smartphone App Preferences for Medication Adherence in Adolescents and Young Adults With Sickle Cell Disease. Pediatr Blood Cancer 2016 May 04;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]
34. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-Centered eHealth Interventions for Children, Adolescents, and Adults With Sickle Cell Disease: Systematic Review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
35. Modi A, Quittner AL. Barriers to treatment adherence for children with cystic fibrosis and asthma: what gets in the way? J Pediatr Psychol 2006 Sep;31(8):846-858. [doi: 10.1093/jpepsy/jsj096] [Medline: 16401680]
36. Ramsey WA, Heidelberg RE, Gilbert AM, Heneghan MB, Badawy SM, Alberts NM. eHealth and mHealth interventions in pediatric cancer: A systematic review of interventions across the cancer continuum. Psychooncology 2020 Jan 19;29(1):17-37. [doi: 10.1002/pon.5280] [Medline: 31692183]
37. Tarricone R, Cucciniello M, Armeni P, Petracca F, Desouza KC, Hall LK, et al. Mobile Health Divide Between Clinicians and Patients in Cancer Care: Results From a Cross-Sectional International Survey. JMIR Mhealth Uhealth 2019 Sep 06;7(9):e13584 [FREE Full text] [doi: 10.2196/13584] [Medline: 31493318]
38. eMarketer. How Many Apps Do Smartphone Owners Use? Most apps are not even retained a full day. 2015. URL: https://emarketer.com/Article/How-Many Apps-Do-Smartphone-Owners-Use/1013309 [accessed 2021-01-26]
39. Hswen Y, Murti V, Vormawor A, Bhattacharjee R, Naslund J. Virtual avatars, gaming, and social media: designing a mobile health app to help children choose healthier food options. J Mob Technol Med 2013;2(2):8-14 [FREE Full text] [doi: 10.7309/jmtm.2.2.3] [Medline: 25419244]
40. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User Preferences and Design Recommendations for an mHealth App to Promote Cystic Fibrosis Self-Management. JMIR Mhealth Uhealth 2014 Oct 24;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
41. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 08;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
42. Frøisland DH, Arsand E, Skårderud F. Improving diabetes care for young people with type 1 diabetes through visual learning on mobile phones: mixed-methods study. J Med Internet Res 2012 Aug 06;14(4):e111 [FREE Full text] [doi: 10.2196/jmir.2155] [Medline: 22868871]
43. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
44. Badawy S, Thompson AA, Kuhns LM. Medication Adherence and Technology-Based Interventions for Adolescents With Chronic Health Conditions: A Few Key Considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
45. Hanghøj S, Boisen KA. Self-reported barriers to medication adherence among chronically ill adolescents: a systematic review. J Adolesc Health 2014 Feb;54(2):121-138. [doi: 10.1016/j.jadohealth.2013.08.009] [Medline: 24182940]
46. Costello R, Foster JM, Grigg J, Eakin MN, Canonica W, Yunus F, Respiratory Effectiveness Group. The Seven Stages of Man: The Role of Developmental Stage on Medication Adherence in Respiratory Diseases. J Allergy Clin Immunol Pract 2016;4(5):813-820 [FREE Full text] [doi: 10.1016/j.jaip.2016.04.002] [Medline: 27587315]
47. Kim MA, Yi J. Life after cancer: how does public stigma increase psychological distress of childhood cancer survivors? Int J Nurs Stud 2014 Dec;51(12):1605-1614. [doi: 10.1016/j.ijnurstu.2014.04.005] [Medline: 24880526]

Abbreviations

6-MP: 6-mercaptopurine
ALL: acute lymphoblastic leukemia
AYAs: adolescents and young adults
BCT: behavior change technique
BCTT: Behavior Change Technique Taxonomy
mHealth: mobile health
NUCATS: Northwestern University Clinical and Translational Sciences

Edited by R Kukafka; submitted 08.10.20; peer-reviewed by A Serlachius; comments to author 04.11.20; revised version received 04.11.20; accepted 07.01.21; published 18.02.21.

Please cite as:
Heneghan MB, Hussain T, Barrera L, Cai SW, Haugen M, Morgan E, Rossoff J, Weinstein J, Hijiya N, Cella D, Badawy SM
Access to Technology and Preferences for an mHealth Intervention to Promote Medication Adherence in Pediatric Acute Lymphoblastic Leukemia: Approach Leveraging Behavior Change Techniques
J Med Internet Res 2021;23(2):e24893
URL: http://www.jmir.org/2021/2/e24893/
doi:10.2196/24893
PMID:33599621

Original Paper

A Web-Based Computer-Tailored Program to Improve Treatment Adherence in Patients With Type 2 Diabetes: Randomized Controlled Trial

Stan Vluggen[1], PhD; Math Candel[2], PhD; Ciska Hoving[1], PhD; Nicolaas C Schaper[3], PhD; Hein de Vries[1], PhD

[1]Department of Health Promotion, Maastricht University, Maastricht, Netherlands

[2]Department of Methodology and Statistics, Maastricht University, Maastricht, Netherlands

[3]Department of Endocrinology and Internal Medicine, Maastricht University Medical Centre, Maastricht, Netherlands

Corresponding Author:
Stan Vluggen, PhD
Department of Health Promotion
Maastricht University
P. Debyeplein 1
Maastricht, 6229 HA
Netherlands
Phone: 31 043 3881557
Email: stan.vluggen@maastrichtuniversity.nl

Abstract

Background: Adherence to core type 2 diabetes mellitus (T2DM) treatment behaviors is suboptimal, and nonadherence is generally not limited to one treatment behavior. The internet holds promise for programs that aim to improve adherence. We developed a computer-tailored eHealth program for patients with T2DM to improve their treatment adherence, that is, adherence to both a healthy lifestyle and medical behaviors.

Objective: The objective of this study is to examine the effectiveness of the eHealth program in a randomized controlled trial.

Methods: Patients with T2DM were recruited by their health professionals and randomized into either the intervention group, that is, access to the eHealth program for 6 months, or a waiting-list control group. In total, 478 participants completed the baseline questionnaire, of which 234 gained access to the eHealth program. Of the 478 participants, 323 were male and 155 were female, the mean age was 60 years, and the participants had unfavorable BMI and HbA_{1c} levels on average. Outcome data were collected through web-based assessments on physical activity (PA) levels, caloric intake from unhealthy snacks, and adherence to oral hypoglycemic agents (OHAs) and insulin therapy. Changes to separate behaviors were standardized and summed into a composite change score representing changes in the overall treatment adherence. Further standardization of this composite change score yielded the primary outcome, which can be interpreted as Cohen *d* (effect size). Standardized change scores observed in separate behaviors acted as secondary outcomes. Mixed linear regression analyses were conducted to examine the effectiveness of the intervention on overall and separate treatment behavior adherence, accommodating relevant covariates and patient nesting.

Results: After the 6-month follow-up assessment, 47.4% (111/234) of participants in the intervention group and 72.5% (177/244) of participants in the control group were retained. The overall treatment adherence improved significantly in the intervention group compared with the control group, reflected by a small effect size (*d*=0.27; 95% CI 0.032 to 0.509; *P*=.03). When considering changes in separate treatment behaviors, a significant decrease was observed only in caloric intake from unhealthy snacks in comparison with the control group (*d*=0.36; 95% CI 0.136 to 0.584; *P*=.002). For adherence to PA (*d*=−0.14; 95% CI −0.388 to 0.109; *P*=.27), OHAs (*d*=0.27; 95% CI −0.027 to 0.457; *P*=.08), and insulin therapy (*d*=0.35; 95% CI −0.066 to 0.773; *P*=.10), no significant changes were observed. These results from the unadjusted analyses were comparable with the results of the adjusted analyses, the per-protocol analyses, and the sensitivity analyses.

Conclusions: Our multibehavior program significantly improved the overall treatment adherence compared with the control group. To further enhance the impact of the intervention in the personal, societal, and economic areas, a wide-scale implementation of our eHealth intervention is suggested.

Trial Registration: Netherlands Trial Register NL664; https://www.trialregister.nl/trial/6664

(J Med Internet Res 2021;23(2):e18524) doi:10.2196/18524

KEYWORDS

type 2 diabetes mellitus; treatment adherence; eHealth; computer-tailoring; randomized controlled trial

Introduction

Globally, 425 million people aged 20 to 79 years live with type 2 diabetes mellitus (T2DM), with expectations of over 600 million people being affected by 2045 [1]. T2DM is associated with considerable morbidity and mortality rates; it reduces patients' quality of life and life expectancy and poses an enormous economic and societal burden [1,2]. Guidelines recommend a series of core treatment behaviors for patients with T2DM. These include healthy lifestyles, that is, improving dietary patterns and increasing physical activity (PA) and, if applicable, adequate adherence to medical strategies such as oral hypoglycemic agents (OHAs) whether or not combined with insulin therapy [3,4]. Despite the chronic progressive nature of T2DM, patients who adhere to these behaviors can live long, high-quality lives [1].

Unfortunately, patients' adherence to separate behaviors is inadequate; dietary and PA targets are not met consistently, and most studies on adherence to medical strategies report adherence prevalence percentages below 80%, which is generally considered insufficient adherence [5-9]. Moreover, less than 5% of the patients diagnosed with T2DM adhere to all treatment behaviors, whereas more than 80% could either improve 2 or more [7]. Poor adherence can result in suboptimal clinical treatment benefits, such as disease worsening, an increase in comorbidity, a reduction in patients' quality of life, increased health care expenditures and hospitalizations, and early mortality [2,10-16].

Nonadherence to core T2DM treatment behaviors such as healthy lifestyles and taking medication is a complex process and a result of an interaction of multiple factors, including social and economic factors, the health care system, characteristics of the disease and therapy, and patient-related factors [17,18]. Although all these factors provide relevant entries for targeting nonadherence, most are difficult to change, and if changed, they may only affect adherence improvements indirectly through patient factors [18]. However, patient-related factors (eg, awareness, beliefs, motivation, self-regulatory capacities) have been shown to be relatively changeable and have a direct impact on treatment adherence [19]. Hence, these determinants need to be addressed in interventions aimed at improving treatment adherence.

Several patient-focused interventions already exist that aim to improve treatment adherence. Most of these interventions pursued improvements in adherence to blood glucose–lowering medication [20-22], of which a minority showed significant improvements in medication adherence and glycemic control. Glycemic control is, however, not only the result of medication adherence but also greatly affected by (un)healthy lifestyle behaviors [18,23]. Therefore, interventions that target a combination of both healthy lifestyle and medical behaviors, that is, a multibehavior approach, might be more likely to be effective [7].

In addition to the multibehavior approach, other factors may enhance the effectiveness of interventions that aim to improve adherence. The internet holds promise for a wide-scale promotion of behavioral change to facilitate the management of T2DM [21,24,25]. Internet interventions as a delivery platform for health promotion and health service activities, also referred to as eHealth, have been shown to be effective, cost-effective, easy to use; have fewer availability restrictions than regular medical consultations; and can temper pressure on health care systems [26-32]. A more advanced eHealth strategy applies computer-tailored technology, an effective strategy that provides patients with tailored content based on unique answers given to a web-based assessment [33,34]. Further success factors of eHealth interventions include the application of a theoretical foundation; provision of interactive tailored content; application of goal-setting strategies and monitoring tools; identification of risk behaviors, using visually supported content; and focusing on distinct behavior change phases, that is, awareness, motivation, and self-regulation [21,24,25,35,36].

However, a recent review on eHealth interventions supporting T2DM management [25] concluded that only one of the 9 included studies reported significant improvements in dietary behavior and PA [36]. Generally, such eHealth interventions often include little interactive content and tailored strategies, are mainly text based, make little use of theoretical foundations and technology, and focus on separate behaviors that play a role in the management of T2DM instead of combining behaviors [21,24,25,35], which may explain the relatively poor results of available interventions.

Hence, eHealth interventions aimed at T2DM treatment adherence might be significantly improved by building on a theoretical base, incorporating computer-tailored technology, providing interactive and visually supported content, and applying a multibehavior approach.

Therefore, we developed an eHealth program for patients with T2DM, including the abovementioned success factors, to improve treatment adherence to core T2DM treatment behaviors, that is, healthy lifestyle and medical behaviors. The main aim of this study is to examine the effectiveness of this program on overall treatment adherence in a randomized controlled trial (RCT). In addition, we examined changes to separate treatment behaviors as a result of the program.

Methods

Study Design

We conducted an RCT including an intervention group and waiting-list control group to examine the effectiveness of a novel eHealth program, *My Diabetes Profile* (MDP), on treatment behavior adherence in patients with T2DM. A more extended description of the program, including its development and content and a trial protocol, is available elsewhere [37]. The study was evaluated and approved by the Medical Ethics Committee of Maastricht University Medical Centre (16-4-171).

The committee concluded that no ethical clearance was needed according to the rules and regulations of the Medical Research Involving Human Subjects Act. The trial is registered in the Netherlands Trial Register (NL6664).

MDP Program

The MDP program aims to improve patient adherence to core T2DM treatment behaviors. This implies improving PA levels; decreasing caloric intake from unhealthy snacks, as this emerged as a major issue in the diet of patients with T2DM in our preliminary work [38]; and increasing adherence to medical strategies, that is, OHAs whether or not combined with insulin therapy [3,4]. A screenshot of the main menu of the MDP program is presented in Multimedia Appendix 1.

The MDP program is theoretically grounded in the Integrated Change Model, which integrates various acknowledged sociocognitive theories that assume a deliberate process when someone engages in (health) behavior [39-43]. The model has frequently been applied to map salient sociocognitive determinants of health behavior and to develop effective web-based computer-tailored interventions aimed at health behavior change accordingly [44,45]. The MDP program is self-guided and facilitated through periodic prompts and reminders to stimulate program engagement and completion. The program provides web-based text and video feedback messages, tailored to determinants and underlying salient beliefs of health behavior change such as knowledge, attitudes, self-efficacy, goal setting, and action planning [43]. The program is divided into 2 nearly identical blocks, each available to users for 3 months. Each block consists of 3 sessions: (1) health risk appraisal; (2) awareness and motivation; and (3) goal setting, action planning, and self-regulation.

The health risk appraisal session provides patients with interactive and tailored content on their risk behaviors. Primarily, adherence levels are assessed for all behaviors the patient was involved in. For behaviors subject to improvement, the participants' intention to change that behavior is assessed. The final part of the first session enables patients to select a single improvable behavior, which will be their focus for the following 3 months while working with the MDP program. In the event of meeting all guideline targets, patients are prompted to select PA as Dutch guidelines recommend any PA beyond the minimum weekly standard of 150 min [46]. A patient who selects a behavior that is accompanied by a low intention to change is navigated to the awareness and motivation session. This second session aims to raise patients' awareness of the need to improve their particular behavior and to increase motivation, with the ultimate purpose of achieving a high intention to change. If a high intention to change is achieved, after either session 1 or session 2, the patient is directed to session 3 on goal setting, action planning, and self-regulation. This session aims to increase the likelihood of a successful translation of the expressed intention into subsequent behavior. This process is facilitated by setting small and realistic goals; forming action plans on where, when, and how to perform the behavior; and forming self-regulation strategies to cope with barriers or situations that may impede adherence.

Participants and Procedure

In the Netherlands, patients usually visit their nurses every 3 months, under the supervision of a physician [3]. Therefore, these nurses were considered to be in an ideal position to recruit patients for this trial. Nationwide, nurses were approached via email, telephone calls, letters to their work address, and social media platforms (eg, LinkedIn and Facebook). They could sign up for the study by contacting the research team directly or by registering via the project website. Nurses were asked to recruit at least eight patients within a period of 6 months. Inclusion criteria for patients were (1) T2DM diagnosis for at least one year, (2) being 40 to 70 years old, (3) using at least one form of oral blood glucose–lowering drugs or insulin, and (4) having no walking disability. Exclusion criteria were (1) not speaking or understanding the Dutch language, (2) having no access to the internet, and (3) using an insulin pump.

After recruitment, nurses filled a brief web-based registration form consisting of the participant's name, telephone number (optional), birth date, most recent HbA_{1c} level (a measure for glycemic control), the year of diabetes diagnosis, current diabetes medical strategy, and email address. Once registered, patients received an email, including log-in data, which primarily provided access to additional study and procedure information before providing informed consent. Participants would then fill the baseline questionnaire after which they were randomly allocated to either the intervention group (receiving program access for 6 months) or the control group (receiving care as usual). Individuals allocated to the control group were informed about their group allocation after baseline completion and notified that they would be invited for the follow-up assessment 6 months later. Moreover, they were informed about the possibility of accessing the MDP program after completing the follow-up assessment as part of the waiting-list control group. Randomization occurred at the individual level by means of computer software randomization. After randomization, nurses were able to review if their patients were allocated to the intervention or control group. For patients who received access to the MDP program, a brief summary of the patient's activity and progress in the program was available to the particular nurse, which could voluntarily be discussed in subsequent face-to-face sessions [37].

Measurements

The baseline questionnaire included 131 questions on demographic characteristics, comorbidities, smoking status, current PA levels, caloric intake from unhealthy snacks, and adherence to OHAs whether or not combined with insulin therapy. The questionnaires were identical for both trial groups.

Demographic Characteristics

Demographics assessed only at baseline included the participant's gender (male or female), age, education level (low: no education up to lower technical education; medium: general secondary education up to secondary vocational education; or high: school of higher general secondary education up to university degree), body length, and nationality. Living arrangement (together or alone), net income (under or above average), work status (salaried or self-employed, no salaried

employment, retired or disabled or incapacitated), T2DM medication type (oral blood glucose–lowering medication, insulin therapy, or a combination), and body weight were assessed at both baseline and follow-up. BMI was calculated as weight per length^2.

Comorbidities

Questions on comorbidity assessed, at baseline only, whether participants were affected in the past or currently have conditions, including depression, stroke, heart failure, myocardial infarction, cancer, chronic obstructive pulmonary disease or asthma or bronchitis, rheumatoid arthritis or osteoarthritis, and Crohn disease.

PA

PA levels were assessed using the validated Short Questionnaire to Assess Health-Enhancing Physical Activity (SQUASH) [47]. SQUASH assesses various domains of PA, for which the average daily hours and minutes, and the number of days per week activities are carried out, are reported. Each domain corresponds to a specific metabolic equivalent of task (MET) value, an intensity and energy expenditure ratio of a task compared with energy expenditure while at rest [48]. As national guideline targets recommend at least moderate PA, that is, ≥3 MET, and because SQUASH includes 2 activities, that is, <3 MET, these activities were excluded [49]. The cumulative number of weekly PA minutes was calculated accordingly.

Unhealthy Snack Intake

Weekly caloric unhealthy snacks intake was assessed using a self-administered food frequency questionnaire (FFQ). The FFQ includes unhealthy snacks identified by earlier studies, complemented with snacks commonly consumed in the Netherlands [50,51]. The unhealthy snacks listed in the FFQ are translated into a particular amount of calories consumed, based on the calorie database of the Dutch nutrition center [52]. A total of weekly caloric intake was calculated based on the participant's intake from unhealthy snacks.

OHA Adherence

Oral drug adherence was measured using the Probabilistic Medication Adherence Scale (ProMAS) questionnaire [53]. The scale includes 18 items that assess a variety of adherence behaviors. To reduce potential recall bias, a period of 3 months was added to every item [54]. This period was chosen because, in the Netherlands, most patients visit their nurse quarterly, and this visit comprises discussing treatment adherence and if pharmacological changes are needed [3]. A sum score was calculated for the 18 items, ranging from 0 to 18, with higher scores representing better adherence.

Insulin Therapy Adherence

Insulin therapy adherence was assessed through an adapted version of the ProMAS questionnaire and included 9 items that were assessed over a 3-month period. Nonrelevant items, that is, items that did not distinguish between adherence and nonadherence to insulin therapy were removed [37]. A sum score was calculated for the 9 items, ranging from 0 to 9, with higher scores representing better adherence.

Primary Outcome, Primary End Point, and Power Calculation

The primary outcome was the composition score of changes in separate treatment behaviors addressed in the program. To create such a composition score, changes in each treatment behavior, that is, changes in PA levels, caloric intake from unhealthy snacks, and OHA and insulin therapy adherence, were standardized into separate change scores. For each treatment behavior in each participant, baseline scores were subtracted from the follow-up scores, yielding a change score. The change score for caloric intake was reversed as the program aimed to decrease caloric intake from unhealthy snacks. To standardize the outcomes of the different behaviors, given the varying units of measurement, the change score of each participant was divided by the pooled SD of the change scores of this specific behavior [55]. The pooled SD of the change scores of both trial groups was used. Finally, per participant, these separate standardized change scores were then summed into a composite change score [56].

The composite change score was transformed, that is, standardized further, to be interpreted as Cohen *d* (effect size), by dividing it for each participant by the pooled SD of these composite change scores. Again, the pooled SD of the composite change scores of both trial groups was used. The standardized composite change score is the primary outcome and Cohen *d* is the primary end point of this study.

The power calculation was based on the primary outcome. We aimed to detect a difference in the mean of the primary end point between trial groups of 0.4 in a two-tailed test at a 5% type I error rate [57]. Considering an intraclass correlation coefficient of 0.02 and a statistical power of 80%, 116 participants per trial group would be sufficient for the trial's follow-up assessment [58]. Given an expected attrition rate of 50%, we aimed to include 464 participants with a completed baseline assessment.

Secondary Outcomes: Changes in Separate Treatment Behaviors

In addition to calculating a standardized composite change score, separate changes in PA levels, caloric intake by unhealthy snacks, and OHA and insulin therapy adherence were calculated. A standardized change score per treatment behavior was calculated per participant by subtracting the baseline score from the follow-up score, yielding a change score. Subsequently, the change score of each participant was divided by the pooled SD of the change scores of this specific behavior, as described above [55]. The difference in the means of these standardized change scores between the intervention and control group can again be interpreted as Cohen *d*, indicating the effect size for separate treatment behaviors [57].

Statistical Analyses

All analyses were conducted using SPSS version 24.0, with a 5% significance level. Frequency and descriptive analyses were used to describe the sample characteristics. Primary and secondary outcomes were analyzed according to the intention-to-treat principle. As participants were nested within nurses participating in the trial, linear mixed regression analyses

were conducted to assess the effectiveness of the MDP program. Covariates in the model were included based on the assumptions of the theoretical framework that was applied in this study, and included gender, age, education level, net income, living arrangement, work status, BMI, HbA_{1c} level, T2DM medication type, recruitment nurse type, and depression status. Results from the unadjusted and adjusted analyses are presented for the primary and secondary outcomes.

Multiple imputation was used for missing values on covariates and outcome variables, which is valid under the assumption that values are missing at random [59,60]. In addition, sensitivity analyses were performed. These consisted of per-protocol analyses and imputation scenarios involving participants of whom the primary outcome was not available. For the per-protocol analyses, results of the unadjusted analyses were presented for the primary and secondary outcomes. For the imputation scenarios, 4 different imputation scenarios were performed for the primary outcome: 2 optimistic and 2 pessimistic scenarios. In the optimistic scenarios, we assumed that, compared with the per-protocol analysis, dropouts improved in the primary outcome, whereas in the pessimistic scenarios, we assumed that they deteriorated. For both the optimistic and pessimistic scenarios, 2 imputation variants were performed, as attrition was unequal in the groups of the trial, and this imbalance might have affected the results of the analysis. In the equal variant, the imputed value was drawn for both the intervention and control groups assuming a normal distribution of the outcome with a mean equal to the condition mean +/−1×SD. In the unequal variant, we used either a mean equal to the intervention mean +/−1×SD for the intervention group or a mean equal to the control mean +/−1.5×SD in the control group. This unequal variant reflects the possibility that the outcome on the average either improved or deteriorated to a lesser extent in the participants of the intervention group, compared with the control group. More effort was required from the participants of the intervention group, which may have affected attrition.

Logistic regression analysis was applied to examine selective attrition after randomization regarding background characteristics, including gender, age, education level, net income, living arrangement, work status, BMI, HbA_{1c} level, T2DM medication type, recruitment nurse type, depression status, and trial condition. In addition, we examined whether participants varied in their baseline adherence to treatment behaviors based on their retention status.

Results

Sample Characteristics

Figure 1 shows the flow of participants throughout the trial from initial recruitment and registration by nurses to the completion of the follow-up assessment and requested program access by control group participants. In total, 669 participants were registered in the program. Overall, 75.9% (508/669) of participants signed the web-based informed consent. Overall, 94.1% (478/508) of participants completed the baseline assessment and were randomly allocated to either the intervention group (234/478, 48.9%) or control group (244/478, 51.0%).

The baseline sample of participants (Table 1) had a mean age of 60.2 (SD 6.78) years, approximately one-third were female, and 40.2% (192/478) were less educated. Most participants lived with their partners. A slight majority of the participants used OHAs as the only medication to control their blood glucose levels, whereas approximately one-third applied a combination of insulin and OHAs. The participants had unfavorable HbA_{1c} levels and BMI on average.

On average, patients could improve their three treatment behaviors. At the first occasion to select a single treatment behavior, most MDP participants chose to improve their PA levels (88/203, 43.3%), followed by decreasing unhealthy snacks intake (66/203, 32.5%), improving OHA adherence (39/203, 19.2%), and improving insulin therapy adherence (10/203, 4.9%). After 3 months, most participants chose to improve PA levels (46/104, 44.2%), followed by decreasing unhealthy snacks intake (39/104, 37.5%), improving OHA adherence (15/104, 14.4%), and improving insulin therapy adherence (4/104, 3.8%). Overall, 41.3% (43/104) of the participants chose the same behavior to improve on at the second occasion as at the start of the program. A total of 3 participants met all guideline targets, either on the first or on the second occasion, and were therefore prompted to select PA, as this was considered improvable regardless of the initial level [37].

The 6-month follow-up assessment was completed by 60.2% (288/478) of participants; 47.4% (111/234) in the intervention group and 72.5% (177/244) in the control group (Figure 1). Control participants were more likely to be retained in the study (OR 2.93, 95% CI 2.001 to 4.283; P<.001). Dropouts did not differ from those who were retained in terms of background characteristics and baseline adherence to treatment behaviors. About one-third of the control group participants requested program access, offered as part of the waiting-list control design, at the end of the study.

Figure 1. Participant and randomization flow throughout the trial.

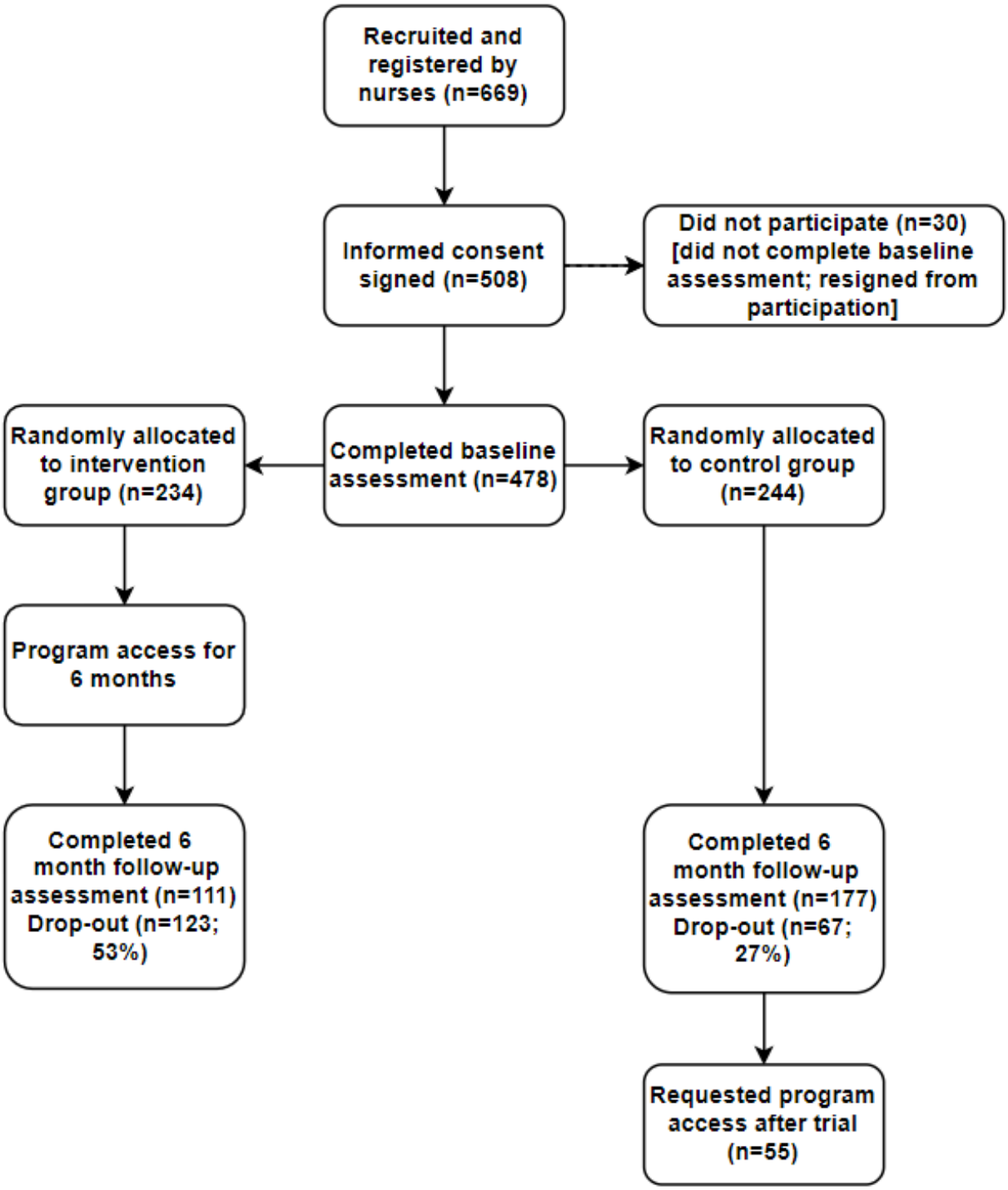


Table 1. Sample characteristics of patients at baseline and comparison of baseline characteristics between the intervention group (n=234) and control group (n=244).

Characteristic	Intervention group (n=234)	Control group (n=244)
Age in years, mean (SD)	60.9 (6.3)	59.4 (7.1)
Gender (female), n (%)	76 (32.5)	79 (32.4)
Education level, n (%)		
Low	90 (38.5)	102 (41.8)
Middle	56 (23.9)	52 (21.3)
High	85 (36.3)	86 (35.2)
Missing data	3 (1.3)	4 (1.9)
Living arrangement, n (%)		
Together with partner	182 (77.8)	188 (77)
Alone	52 (22.2)	55 (22.5)
Missing data	N/A[a]	1 (0.4)
Work status, n (%)		
Salaried or self-employed	96 (41.0)	111 (45.5)
No salaried employment	35 (15.0)	39 (16.0)
Retired	75 (32.0)	70 (28.7)
Disabled or incapacitated	28 (12.0)	24 (9.8)
Net income, n (%)		
Under-average income	60 (25.6)	70 (28.7)
Above-average income	116 (49.6)	122 (50.0)
Missing data	58 (24.8)	52 (21.3)
Diabetes medication type, n (%)		
OHA[b] only	143 (61.1)	151 (61.9)
Insulin therapy only	11 (4.7)	21 (8.6)
OHA and insulin therapy	80 (34.2)	72 (29.5)
Recruitment nurse, n (%)		
Practice nurse	165 (70.5)	177 (72.5)
Diabetes nurse	69 (29.5)	67 (27.5)
Depression status, n (%)		
Never or in the past	224 (95.7)	228 (93.4)
Current	10 (4.3)	16 (6.6)
HbA_{1c}[c] (mmol/mol), mean (SD)[d]	56.6 (11.8)	57.1 (11.8)
BMI (kg/m^2), mean (SD)	30.8 (4.9)	31.2 (5.1)
OHA adherence score, mean (SD)	13.5 (3.5)	12.9 (4.0)
IT[e] adherence score, mean (SD)	7.4 (1.8)	7.9 (1.6)
Physical activity (min), mean (SD)[f]	868 (1031)	764 (796)
Snack intake (cal), mean (SD)[f]	1746 (1435)	1676 (1375)

[a]N/A: not applicable.

[b]OHA: oral hypoglycemic agent.

[c]HbA_{1c}: glycosylated hemoglobin.

[d]This equals an HbA_{1c} of 7.4%.

[e]IT: insulin therapy.

[f]Average number of weekly minutes or calories.

Effect Analyses on Primary Outcome (Overall Treatment Adherence)

The results of the unadjusted and adjusted linear mixed regression analyses for the primary outcome are shown in Table 2. The result of the unadjusted analysis shows that allocation to the MDP program had a significant small effect on overall treatment adherence (*d*=0.27; 95% CI 0.032 to 0.509; *P*=.03). The adjusted result showed a similar small effect (*d*=0.25; 95% CI 0.010 to 0.495; *P*=.04). In total, 77.5% (86/111) of the participants in the intervention group showed improvement in overall treatment adherence compared with 60.5% (107/177) in the control group.

Table 2. Results of linear mixed regression analysis on multiple imputed data sets: unadjusted and adjusted model.

Linear mixed regression analysis	Regression coefficient (Cohen *d*)	95% CI	*t* test (*df*)	*P* value
Unadjusted model (trial group)				
Intervention group	.27	0.032 to 0.509	2.229 (306.60)	.03
Control group[a]	N/A[b]	N/A	N/A	N/A
Adjusted model (trial group)				
Intervention group	0.25	0.010 to 0.495	2.05 (302.34)	.04
Control group[a]	N/A	N/A	N/A	N/A
BMI	−0.01	−0.032 to 0.014	−0.78 (558.01)	.43
Recruitment nurse				
Practice nurse	−0.05	−0.370 to 0.271	−0.31 (223.18)	.76
Diabetes nurse[a]	N/A	N/A	N/A	N/A
Diabetes medication type				
OHA[c] only	0.01	−0.182 to 0.375	0.68 (454.77)	.50
Insulin therapy only	−0.06	−0.572 to 0.451	−0.23 (288.59)	.82
OHA and insulin therapy[a]	N/A	N/A	N/A	N/A
HbA_{1c}[d]	0.01	−0.005 to 0.019	1.19 (195.41)	.23
Depression status				
Never or in the past	0.12	−0.650 to 0.407	−0.45 (307.75)	.65
Current[a]	N/A	N/A	N/A	N/A
Gender				
Male	−0.10	−0.381 to 0.174	−0.73 (259.33)	.46
Female[a]	N/A	N/A	N/A	N/A
Age (years)	−0.01	−0.030 to 0.016	−0.59 (354.14)	.56
Work status				
Salaried or self-employed	−0.41	0.818 to 0.004	−1.95 (321.68)	.05
No salaried employment	−0.56	−1.049 to −0.077	−2.28 (277.27)	.02
Retired	−0.28	−0.733 to 0.167	−1.24 (319.11)	.22
Disabled or incapacitated[a]	N/A	N/A	N/A	N/A
Living arrangement				
Together with a partner	0.22	−0.082 to 0.522	−1.43 (291.29)	.15
Alone[a]	N/A	N/A	N/A	N/A
Education level				
Low	−0.37	−0.668 to −0.069	−2.42 (261.03)	.02
Middle	−0.01	−0.308 to 0.288	−0.07 (602.33)	.95
High[a]	N/A	N/A	N/A	N/A
Net income				
Under-average income	0.28	−0.095 to 0.659	1.48 (147.14)	.14
Above-average income[a]	N/A	N/A	N/A	N/A

[a]Reference category.

[b]N/A: not applicable.

[c]OHA: oral hypoglycemic agent.

[d]HbA_{1c}: glycosylated hemoglobin.

Effect Analyses on Secondary Outcomes (Separate Behavior Adherence)

The results of unadjusted linear mixed regression analyses for the secondary outcomes are shown in Table 3; the results of the adjusted analyses are presented in Multimedia Appendix 2. The results of the unadjusted analyses show that allocation to the MDP program had a significant small-to-medium effect on the decrease in caloric intake from unhealthy snacks (*d*=0.36; 95% CI 0.136 to 0.584; *P*=.002). The effects of oral hypoglycemic adherence (*d*=0.22; 95% CI −0.027 to 0.457; *P*=.08) and insulin therapy adherence (*d*=0.35; 95% CI −0.066 to 0.773; *P*=.10) were small but not significant. No effect was observed for PA (*d*=−0.14; 95% CI −0.388 to 0.109; *P*=.27). The results of the adjusted analyses of the secondary outcomes remained roughly equal to those of the unadjusted analyses.

Table 3. Results of the linear mixed regression analyses for separate treatment behaviors on multiple imputed data sets: unadjusted models.

Unadjusted models	Regression coefficient (Cohen *d*)	95% CI	*t* test (*df*)	*P* value
OHA[a] score				
Intervention group	0.22	−0.027 to 0.457	1.748 (338.47)	.08
Control group[b]	N/A[c]	N/A	N/A	N/A
IT[d] score				
Intervention group	0.350.35	−0.066 to 0.773	1.658 (278.49)	.10
Control group[b]	N/A	N/A	N/A	N/A
PA[e] level				
Intervention group	−0.14	−0.388 to 0.109	−1.107 (235.32)	.27
Control group[b]	N/A	N/A	N/A	N/A
Snack intake				
Intervention group	0.36	0.136 to 0.584	3.150 (474.10)	.002
Control group[b]	N/A	N/A	N/A	N/A

[a]OHA: oral hypoglycemic agent.

[b]Reference category.

[c]N/A: not applicable.

[d]IT: insulin therapy.

[e]PA: physical activity.

Sensitivity Analyses

Table 4 shows the adherence scores at baseline and follow-up per trial group and the results of the unadjusted per-protocol analyses. The effects on the primary and secondary outcomes were comparable with the intention-to-treat analyses. The results of the optimistic and pessimistic sensitivity analyses for the primary outcome partially reflected the results of the intention-to-treat and per-protocol analyses. After replicating the unadjusted linear mixed regression analyses following an equal and unequal optimistic imputation scenario, the intervention effect remained significant both under the equal (*d*=0.39; 95% CI 0.201 to 0.579; *P*<.001) and the unequal imputation (*d*=0.33; 95% CI 0.120 to 0.537; *P*=.002). Following an equal and unequal pessimistic imputation scenario, the intervention effect became nonsignificant under the equal (*d*=−0.13; 95 CI −0.353 to 0.090; *P*=.25) and unequal imputation (*d*=−0.04; 95% CI −0.271 to 0.189; *P*=.73).

In the intervention group, 73.9% (82/111) of the participants reduced their intake of unhealthy snacks compared with 54.2% (96/177) in the control group. One participant in the control group did not change in unhealthy snack score. With regard to oral blood glucose–lowering medication, 50.5% (52/103) of the intervention group participants improved their adherence compared with 42.9% (70/163) in the control group. 24.3% (25/103) of the participants in the intervention group did not change in OHA score compared with 20.9% (34/163) in the control group. With regard to insulin therapy adherence, 30.2% (13/43) of the participants in the intervention group improved their adherence compared with 27.9% (19/68) in the control group. Furthermore, 58.1% (25/43) of the participants in the intervention group did not change in insulin therapy score compared with 55.9% (38/68) in the control group.

Table 4. Adherence scores per trial group at baseline and follow-up, and the unadjusted per-protocol analyses for primary and secondary outcomes.

Treatment behavior	Adherence scores in the intervention group, mean (SD)			Adherence scores in the control group, mean (SD)			Cohen *d* (95% CI)	*t* test (*df*)	*P* value
	Baseline	Follow-up	n (%)	Baseline	Follow-up	n (%)			
Overall adherence	N/A[a]	N/A	111 (47.4)	N/A	N/A	177 (72.5)	0.24 (0.005 to 0.481)	2.006 (286)	.046
OHA[b] score	13.4 (3.4)	14.3 (3.6)	103 (44.0)	13.1 (3.8)	13.4 (3.6)	163 (66.8)	0.18 (−0.070 to 0.423)	1.407 (264)	.16
IT[c] score	7.4 (1.8)	8.0 (1.6)	43 (18.4)	7.7 (1.7)	7.9 (1.7)	68 (27.9)	0.28 (−0.102 to 0.658)	1.453 (109)	.15
PA[d] level[e]	865 (1141)	833 (741)	111 (47.4)	789 (769)	884 (777)	177 (72.5)	−0.07 (−0.265 to 0.130)	−0.675 (286)	.50
Snack intake[f]	1857 (1330)	1269 (1182)	111 (47.4)	1656 (1331)	1496 (1108)	177 (72.5)	0.38 (0.098 to 0.661)	2.655 (286)	.009

[a]N/A: not applicable.

[b]OHA: oral hypoglycemic agent.

[c]IT: insulin therapy.

[d]PA: physical activity.

[e]Average number of minutes per week.

[f]Average number of calories per week.

Discussion

Principal Findings and Comparison With Previous Work

This study examined the effectiveness of a novel web-based, computer-tailored program, *MDP*, on overall adherence to core treatment behaviors in patients with T2DM. In addition, we explored the effects of the MDP program on each separate behavior. The MDP program improved overall adherence with a small, significant effect size. With regard to changes in adherence to each separate behavior, a small-to-medium significant effect size was observed for decreasing caloric intake from unhealthy snacks. Despite observing small-to-medium effect sizes for medication-taking behaviors, these showed no significance. With regard to adherence to PA, no significant changes were observed.

Our study focuses on improving adherence to multiple treatment behaviors in T2DM and is, to our knowledge, the first to subsequently quantify program effects in terms of the overall effect size. Generally, intervention studies put limited emphasis on exploring the overall change across risk behaviors [56]. Most studies focused on improving separate treatment behaviors, and the few targeting multiple risk behaviors have mainly examined changes in separate risk behaviors [24,25]. When risk behaviors co-occur, which is the case in the vast majority of patients with T2DM, the adverse effects on health and health outcomes are the largest [7,55,56]. Evaluating the effect of multibehavior interventions requires methods to quantify changes across several behaviors. Prochaska et al [55] supported quantifying the overall change in multiple behaviors by first calculating standardized change scores of each separate behavior and subsequently adding these scores [56,61]. However, although the application of such a composite change score has considerable advantages, it may be difficult to interpret as it is an abstract number. After further transforming this score, treatment effects can be interpreted as effect sizes (Cohen *d*), thereby increasing the interpretability of the observed results. Such an effect size also allows for comparison between distinct multibehavior interventions and examines the overall impact on health behavior change, which is not possible when focusing on changes in separate behaviors [56,62]. We recommend future studies to report multibehavior intervention effects in terms of effect sizes to improve interpretability and allow comparisons across studies.

Findings from this study appear to be robust and credible because the results of the unadjusted and adjusted intention-to-treat analyses were comparable with the per-protocol and sensitivity analyses. However, in the pessimistic missing data imputation scenarios, the main effect of the intervention became nonsignificant. Probably, the complete case scenario is the most accurate reflection of the actual intervention effect in this case because the attrition analysis revealed that differential attrition with regard to the demographic characteristics of the participants was absent [63-65].

When inspecting the nature of the overall effect and looking at the individual behaviors, we observed small-to-medium effect sizes for caloric intake from unhealthy snacks (d=0.36), insulin therapy adherence (d=0.35), OHA adherence (d=0.22), and a negligible effect size for PA (d=−0.14). Of these effects, only the effect for caloric intake was significant. Our results are in line with an earlier RCT in patients with T2DM on the effects of a web-based diabetes support program [37]. Directly after a 4-month intervention, significant improvements in healthy eating (d=0.32), fat intake (d=0.28), and PA (d=0.19) were observed in this trial; however, a negligible effect was observed for medication taking [36]. The effects in both our study and the aforementioned study may be related to specific success factors of eHealth interventions. Both interventions incorporated a theoretical foundation, interactive tailored content, and

addressed multiple behaviors involved in the treatment of T2DM. Although it is difficult to examine the exact effect of such success factors, our findings support the conclusion that interventions applying sound theoretical motivational theories as a basis, interactive tailored content, and a multibehavior approach can have relevant effects [21,24,25,35,36].

A significant effect of our program was seen on eating behavior, that is, a decrease in caloric intake from unhealthy snacks. Improving dietary patterns in patients with T2DM is multifaceted, and generally, a decrease in fat intake is strived for [30,66]. Intake of unhealthy snacks recently emerged as a major novel issue in diets of patients with T2DM, according to both health professionals and patients themselves [38], and about a third of our participants selected this behavior to improve in the MDP program. Perhaps, the novelty of this diet topic and the detailed health risk appraisal complemented with personal feedback may have informed and alerted patients in such a way that they were triggered to successfully pursue a decrease in their caloric intake via snacks [67]. However, more research is needed to examine why a significant effect of our program occurred in caloric intake from unhealthy snacks.

No significant effects were observed for OHA and insulin therapy adherence and for PA. Observing no effects on PA levels seems to be common in digital multibehavior intervention studies [31,45,61,68-71]. However, PA was the behavior most often chosen by patients, and they were provided with a detailed health risk appraisal on their current PA levels. A more detailed analysis revealed that these patients had a high willingness to increase their PA and almost all could have chosen other topics to improve; therefore, they were not *forced* to pursue PA improvements. There are several potential explanations for the lack of any effect on the PA of our program. Patients could have overestimated their need and willingness to improve PA levels, as on average in this group, guideline targets were six-fold higher than the target of 150 min PA per week [72]. Note that these were self-reported values and were probably an overestimation, as several studies have shown that many persons are unable to reliably estimate their PA levels [73]. Using modern technology, such as accelerometers, this barrier might be overcome in the future [72]. It may also be that patients chose PA, as opposed to medication adherence improvements, as patients considered healthy lifestyle domains as more crucial to their health than medication taking [74,75]. Although improvements emerged in the other domains, that is, intake of unhealthy snacks, it might have been that improvements in this behavior required considerable self-control efforts. In turn, self-control spent on decreasing caloric intake may have depleted resources for further volitional efforts, such as improving or maintaining already high PA levels [76]. In fact, recent studies indeed show that high levels of self-control are required to translate short-term intentions into pursued PA improvements [77]. However, further research is needed to determine why multibehavior internet interventions seem to have such a limited effect on PA levels. For instance, reasons for failure could be explored in-depth using qualitative interviews, to analyze whether specific and effective action plans were made and whether PA plans were combined with other adherence activities that could have led to overdepletion in certain patients [76,78].

Strengths and Limitations

Primarily, the multibehavior approach to improve treatment adherence is a strength of our study. Existing interventions have largely focused on improving single behaviors, whereas the management of T2DM is multifaceted and treatment nonadherence co-occurs across treatment behaviors. Second, the program was theory based and applied previously identified success factors for effective web-based self-management programs. Third, our nationwide recruitment and quite robust findings enhance the generalizability and credibility of our results. In fact, in a large Dutch survey study investigating characteristics of Dutch patients with T2DM, the distribution of education level, average age, living arrangement, paid employment, HbA_{1c} level, and BMI were comparable with the sample characteristics in our study [79]. In our study, slightly more people used T2DM medications, as this was a study requirement. Fourth, our study improved interpretability of the analysis results on the mean difference between intervention and control by transforming it into an effect size, which may also simplify comparisons of results of similar interventions in meta-analyses [57]. Moreover, changes in all behaviors were incorporated in the primary outcome, including changes to those behaviors that patients did not choose to receive feedback on in the intervention. Although this could potentially have reduced the effect identified, the overall score allowed us to correct our findings for changes in other behaviors (ie, improvements or compensation trade-offs) [74,80]. Finally, the attrition rate in our study was 53% in the intervention group and 27.5% in the control group. We did not observe differences between the participants who dropped out and those who did not; however, we cannot exclude that this might have affected our results. Moreover, the higher dropout rate in the intervention group could perhaps be a consequence of the necessary time investment. However, these results seem favorable as attrition rates reaching 60% to 80% are common in web-based interventions [81-83]; however, use of the program could probably be further stimulated by integrating it more in daily care. Feedback from noncompleters could yield valuable input to explore and improve retention rates and should therefore be addressed in future trials.

The limitations of our study are mainly methodological in nature. First, adherence data were collected through self-report questionnaires, which can be prone to social desirability issues and behavior overestimation [73,84]. To reduce the social desirability issue, future trials should consider objective measurement instruments for behaviors where this is possible, for example, accelerometers for PA and electronic monitor systems for medication adherence. However, the impact of such biases within this trial was likely reduced by applying baseline follow-up change scores and identical adherence assessments for participants in both trial groups. Second, we did not assess clinical outcomes such as glycemic control. However, as we did not include a postintervention follow-up period, it may be unlikely that such outcomes will be observed immediately after our intervention. Third, the nurses in our study were not blinded. For intervention group participants, nurses could voluntarily

review and discuss the patient's activity and progress in the program. However, it should be noted that nurses were not involved in the outcome assessment. Fourth, despite the advantages of a waiting-list control design that allows control participants to access the intervention after trial completion, some limitations apply such as a potential overestimation of treatment effects and delayed change by control group participants [85]. In our trial, however, the control group improved marginally in all outcomes, which may indicate a Hawthorne effect—the awareness of participants of being studied and the possible effect on behavior as a result—which is in accordance with trials not employing a waiting-list design [86]. Finally, we did not assess the long-term effects of our MDP program.

Conclusions and Implications

The MDP program yielded larger improvements in overall treatment adherence postintervention, compared with our control group, reflected by a small overall effect size. Changes in separate behaviors yielded a significant small-to-medium effect size for decreasing caloric intake from unhealthy snacks, whereas small-to-medium but statistically nonsignificant effects were observed for medication-taking behaviors. Small-to-medium effect sizes, as observed in our study, may be of importance when multiplied to the population level [36], as the impact of the program depends not only on its effectiveness but also on its reach. To increase the reach of the MDP program, dissemination challenges could be explored, for example, if the current and scalable recruitment strategy is feasible in practice. Researchers could investigate health professionals' willingness to adopt and implement the MDP program and whether reviewing the patient's activity in the program is of added value to the professional. Implications for research include conducting a cost-effectiveness evaluation and a process evaluation. A process evaluation could yield information on the appreciation of the program, its working mechanisms, and adherence to the intervention and provide insights into the reasons for dropping out of the program. In addition, research could investigate long-term intervention effects and effects of the program on biomedical and societal outcomes such as glycemic control and quality of life. Further research could investigate the need for a more refined, composite score that may address the relative importance of different treatment elements in improving T2DM management. Finally, more research is needed to investigate how PA levels could be improved or maintained through eHealth interventions that aim for adherence improvements in multiple T2DM behaviors.

Acknowledgments

This work was supported by the MUMC+ (Maastricht University Medical Centre) Strategy Horizon 2020. The funding source had no involvement in preparing or conducting the research or research article.

Authors' Contributions

SV, CH, NS, and HV designed the study. SV conducted the study and statistical analyses and prepared the manuscript. MC guided the statistical analyses and performed the sample size calculation. All authors reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Main menu of the My Diabetes Profile program.
[PNG File , 406 KB - jmir_v23i2e18524_app1.png]

Multimedia Appendix 2
Adjusted secondary outcomes of the models after multiple imputations.
[DOCX File , 23 KB - jmir_v23i2e18524_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1187 KB - jmir_v23i2e18524_app3.pdf]

References

1. IDF Diabetes Atlas 8th Edition. International Diabetes Federation. 2017. URL: https://www.diabetesatlas.org/upload/resources/previous/files/8/IDF_DA_8e-EN-final.pdf [accessed 2019-03-20]
2. van't Riet E, Schram MT, Abbink EJ, Admiraal WM, Dijk-Schaap MW, Holleman F, et al. The diabetes pearl: diabetes biobanking in the Netherlands. BMC Public Health 2012 Nov 6;12:949 [FREE Full text] [doi: 10.1186/1471-2458-12-949] [Medline: 23130988]
3. De NDF Zorgstandaard diabetes Type 2 volwassenen 2013. Nederlandse Diabetes Federatie. URL: http://www.zorgstandaarddiabetes.nl/type-2/ [accessed 2017-10-12]

4. García-Pérez LE, Alvarez M, Dilla T, Gil-Guillén V, Orozco-Beltrán D. Adherence to therapies in patients with type 2 diabetes. Diabetes Ther 2013 Dec;4(2):175-194 [FREE Full text] [doi: 10.1007/s13300-013-0034-y] [Medline: 23990497]
5. Nelson KM, Reiber G, Boyko EJ, NHANES III. Diet and exercise among adults with type 2 diabetes: findings from the third national health and nutrition examination survey (NHANES III). Diabetes Care 2002 Oct;25(10):1722-1728. [doi: 10.2337/diacare.25.10.1722] [Medline: 12351468]
6. Morrato EH, Hill JO, Wyatt HR, Ghushchyan V, Sullivan PW. Physical activity in US adults with diabetes and at risk for developing diabetes, 2003. Diabetes Care 2007 Feb;30(2):203-209. [doi: 10.2337/dc06-1128] [Medline: 17259482]
7. King DE, Mainous AG, Carnemolla M, Everett CJ. Adherence to healthy lifestyle habits in US adults, 1988-2006. Am J Med 2009 Jun;122(6):528-534. [doi: 10.1016/j.amjmed.2008.11.013] [Medline: 19486715]
8. Krass I, Schieback P, Dhippayom T. Adherence to diabetes medication: a systematic review. Diabet Med 2015 Jun 9;32(6):725-737. [doi: 10.1111/dme.12651] [Medline: 25440507]
9. Katz LL, Anderson BJ, McKay SV, Izquierdo R, Casey TL, Higgins LA, TODAY Study Group. Correlates of medication adherence in the today cohort of youth with type 2 diabetes. Diabetes Care 2016 Nov 28;39(11):1956-1962 [FREE Full text] [doi: 10.2337/dc15-2296] [Medline: 27352955]
10. Funnell MM, Anderson RM. Empowerment and self-management of diabetes. Clinical Diabetes 2004 Jul 1;22(3):123-127. [doi: 10.2337/diaclin.22.3.123]
11. Asche C, LaFleur J, Conner C. A review of diabetes treatment adherence and the association with clinical and economic outcomes. Clin Ther 2011 Jan;33(1):74-109. [doi: 10.1016/j.clinthera.2011.01.019] [Medline: 21397776]
12. Ho PM, Rumsfeld JS, Masoudi FA, McClure DL, Plomondon ME, Steiner JF, et al. Effect of medication nonadherence on hospitalization and mortality among patients with diabetes mellitus. Arch Intern Med 2006 Sep 25;166(17):1836-1841. [doi: 10.1001/archinte.166.17.1836] [Medline: 17000939]
13. Lorber D. Importance of cardiovascular disease risk management in patients with type 2 diabetes mellitus. DMSO 2014 May:169. [doi: 10.2147/dmso.s61438]
14. Barr EL, Zimmet PZ, Welborn TA, Jolley D, Magliano DJ, Dunstan DW, et al. Risk of cardiovascular and all-cause mortality in individuals with diabetes mellitus, impaired fasting glucose, and impaired glucose tolerance. Circulation 2007 Jul 10;116(2):151-157. [doi: 10.1161/circulationaha.106.685628]
15. Stratton I, Adler A, Neil H, Matthews D, Manley S, Cull C, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. Br Med J 2000 Aug 12;321(7258):405-412 [FREE Full text] [doi: 10.1136/bmj.321.7258.405] [Medline: 10938048]
16. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. N Engl J Med 2008 Oct 9;359(15):1577-1589. [doi: 10.1056/nejmoa0806470]
17. Jaam M, Awaisu A, Ibrahim MI, Kheir N. Synthesizing and appraising the quality of the evidence on factors associated with medication adherence in diabetes: a systematic review of systematic reviews. Value Health Reg Issues 2017 Sep;13:82-91. [doi: 10.1016/j.vhri.2017.09.001] [Medline: 29073997]
18. Sabaté E. Adherence to Long-term Therapies Evidence for Action. World Health Organization. 2003. URL: https://apps.who.int/iris/bitstream/handle/10665/42682/9241545992.pdf [accessed 2018-09-02]
19. French DP, Wade AN, Farmer AJ. Predicting self-care behaviours of patients with type 2 diabetes: the importance of beliefs about behaviour, not just beliefs about illness. J Psychosom Res 2013 Apr;74(4):327-333. [doi: 10.1016/j.jpsychores.2012.12.008] [Medline: 23497835]
20. Vermeire E, Wens J, Van Royen P, Biot Y, Hearnshaw H, Lindenmeyer A. Interventions for improving adherence to treatment recommendations in people with type 2 diabetes mellitus. The Cochrane database of systematic reviews 2005 Apr 20:3638. [doi: 10.1002/14651858.CD003638]
21. Williams JLS, Walker RJ, Smalls BL, Campbell JA, Egede LE. Effective interventions to improve medication adherence in type 2 diabetes: a systematic review. Diabetes Manag (Lond) 2014 Jan 1;4(1):29-48 [FREE Full text] [doi: 10.2217/dmt.13.62] [Medline: 25214893]
22. Sapkota S, Brien J, Greenfield J, Aslani P. A systematic review of interventions addressing adherence to anti-diabetic medications in patients with type 2 diabetes--impact on adherence. PLoS One 2015 Feb 24;10(2):e0118296 [FREE Full text] [doi: 10.1371/journal.pone.0118296] [Medline: 25710465]
23. Fan L, Sidani S. Effectiveness of Diabetes Self-management Education Intervention Elements: A Meta-analysis. Canadian Journal of Diabetes 2009 Jan;33(1):18-26. [doi: 10.1016/s1499-2671(09)31005-9]
24. Ramadas A, Quek K, Chan C, Oldenburg B. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. Int J Med Inform 2011 Jun;80(6):389-405. [doi: 10.1016/j.ijmedinf.2011.02.002] [Medline: 21481632]
25. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014 Mar;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
26. de Vries H, Kremers SPJ, Smeets T, Brug J, Eijmael K. The effectiveness of tailored feedback and action plans in an intervention addressing multiple health behaviors. Am J Health Promot 2008 Jul;22(6):417-425. [doi: 10.4278/ajhp.22.6.417] [Medline: 18677882]

27. Pal K, Eastwood S, Michie S, Farmer A, Barnard M, Peacock R. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. The Cochrane database of systematic reviews 2013 Mar 28:8776. [doi: 10.1002/14651858.cd008776]
28. Smit ES, Evers SM, de Vries H, Hoving C. Cost-effectiveness and cost-utility of Internet-based computer tailoring for smoking cessation. J Med Internet Res 2013 Mar 14;15(3):e57 [FREE Full text] [doi: 10.2196/jmir.2059] [Medline: 23491820]
29. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
30. Clark M, Hampson S, Avery L, Simpson R. Effects of a tailored lifestyle self-management intervention in patients with type 2 diabetes. Br J Health Psychol 2004 Sep;9(Pt 3):365-379. [doi: 10.1348/1359107041557066] [Medline: 15296683]
31. Boudreau F, Godin G, Poirier P. Effectiveness of a computer-tailored print-based physical activity intervention among French Canadians with type 2 diabetes in a real-life setting. Health Educ Res 2011 Mar 16;26(4):573-585. [doi: 10.1093/her/cyr008]
32. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun 18;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
33. de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. Patient Educ Couns 1999 Feb;36(2):99-105. [doi: 10.1016/s0738-3991(98)00127-x] [Medline: 10223015]
34. Kassavou A, Mirzaei V, Brimicombe J, Edwards S, Massou E, Prevost AT, et al. A highly tailored text and voice messaging intervention to improve medication adherence in patients with either or both hypertension and type 2 diabetes in a UK primary care setting: feasibility randomized controlled trial of clinical effectiveness. J Med Internet Res 2020 May 19;22(5):e16629 [FREE Full text] [doi: 10.2196/16629] [Medline: 32427113]
35. Otten W, Rövekamp T, Keer M, de Graaf A, van Keulen HM. Mogelijkheden Voor Webbased Ondersteuning Bij Zelfmanagement Van Type 2 Diabetes Mellitus. TNO. 2015. URL: https://repository.tno.nl/islandora/object/uuid%3Ab49d1c98-362a-4ae5-857e-0b8c9779a9e3 [accessed 2018-10-14]
36. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Outcomes of minimal and moderate support versions of an internet-based diabetes self-management support program. J Gen Intern Med 2010 Dec 17;25(12):1315-1322 [FREE Full text] [doi: 10.1007/s11606-010-1480-0] [Medline: 20714820]
37. Vluggen S, Hoving C, Schaper N, de Vries H. A web-based program to improve treatment adherence in patients with type 2 diabetes: development and study protocol. Contemp Clin Trials 2018 Nov;74:38-45. [doi: 10.1016/j.cct.2018.10.002] [Medline: 30290275]
38. Vluggen S, Hoving C, Schaper N, de Vries H. Exploring beliefs on diabetes treatment adherence among Dutch type 2 diabetes patients and healthcare providers. Patient Educ Couns 2018 Jan;101(1):92-98. [doi: 10.1016/j.pec.2017.07.009] [Medline: 28729129]
39. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
40. Janz NK, Becker MH. The Health Belief Model: A Decade Later. Health Education Quarterly 2016 Sep 04;11(1):1-47. [doi: 10.1177/109019818401100101]
41. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: Toward an integrative model of change. Journal of Consulting and Clinical Psychology 1983;51(3):390-395. [doi: 10.1037/0022-006x.51.3.390]
42. Sniehotta FF, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]
43. de Vries H. An integrated approach for understanding health behavior; the I-change model as an example. PBSIJ 2017 Mar 9;2(2):1-6. [doi: 10.19080/pbsij.2017.02.555585]
44. Walthouwer MJ, Oenema A, Lechner L, de Vries H. Comparing a video and text version of a web-based computer-tailored intervention for obesity prevention: a randomized controlled trial. J Med Internet Res 2015 Oct 19;17(10):e236 [FREE Full text] [doi: 10.2196/jmir.4083] [Medline: 26481772]
45. Schulz DN, Kremers SP, Vandelanotte C, van Adrichem MJ, Schneider F, Candel MJ, et al. Effects of a web-based tailored multiple-lifestyle intervention for adults: a two-year randomized controlled trial comparing sequential and simultaneous delivery modes. J Med Internet Res 2014 Jan 27;16(1):e26 [FREE Full text] [doi: 10.2196/jmir.3094] [Medline: 24472854]
46. Beweegrichtlijnen 2017. Gezondheidsraad. 2017 Aug. URL: https://www.rijksoverheid.nl/documenten/rapporten/2018/11/12/beweegrichtlijnen-2017 [accessed 2019-02-03]
47. Wendel-Vos G. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiology 2003 Dec;56(12):1163-1169. [doi: 10.1016/s0895-4356(03)00220-8]
48. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S498-S504. [doi: 10.1097/00005768-200009001-00009] [Medline: 10993420]

49. Weggemans RM, Backx FJ, Borghouts L, Chinapaw M, Hopman MT, Koster A, Committee Dutch Physical Activity Guidelines 2017. The 2017 dutch physical activity guidelines. Int J Behav Nutr Phys Act 2018 Jun 25;15(1):58 [FREE Full text] [doi: 10.1186/s12966-018-0661-9] [Medline: 29940977]
50. Voedingscentrum. Omgaan met producten buiten de schijf van vijf. Den Haag: Voedingscentrum URL: http://www.voedingscentrum.nl/nl/gezond-eten-met-de-schijf-van-vijf/omgaan-met-producten-buiten-de-schijf-van-vijf.aspx [accessed 2017-12-07]
51. Verhoeven A, Adriaanse M, Evers C, de Ridder D. The power of habits: unhealthy snacking behaviour is primarily predicted by habit strength. Br J Health Psychol 2012 Nov;17(4):758-770. [doi: 10.1111/j.2044-8287.2012.02070.x] [Medline: 22385098]
52. Voedingscentrum. Hoeveel calorieën zitten erin? - Caloriechecker. Den Haag: Voedingscentrum URL: https://www.voedingscentrum.nl/nl/service/vraag-en-antwoord/gezonde-voeding-en-voedingsstoffen/hoeveel-calorieen-zitten-erin-.aspx [accessed 2019-02-18]
53. Kleppe M, Lacroix J, Ham J, Midden C. The development of the ProMAS: a Probabilistic Medication Adherence Scale. PPA 2015 Mar;9:355-367. [doi: 10.2147/ppa.s76749]
54. Clifford S, Perez-Nieves M, Skalicky AM, Reaney M, Coyne KS. A systematic literature review of methodologies used to assess medication adherence in patients with diabetes. Curr Med Res Opin 2014 Jun 11;30(6):1071-1085. [doi: 10.1185/03007995.2014.884491] [Medline: 24432796]
55. Prochaska JJ, Velicer WF, Nigg CR, Prochaska JO. Methods of quantifying change in multiple risk factor interventions. Prev Med 2008 Mar;46(3):260-265 [FREE Full text] [doi: 10.1016/j.ypmed.2007.07.035] [Medline: 18319099]
56. Drake BF, Quintiliani LM, Sapp AL, Li Y, Harley AE, Emmons KM, et al. Comparing strategies to assess multiple behavior change in behavioral intervention studies. Transl Behav Med 2013 Mar 1;3(1):114-121 [FREE Full text] [doi: 10.1007/s13142-013-0195-2] [Medline: 23504621]
57. Sullivan G, Feinn R. Using effect size-or why the P value is not enough. J Grad Med Educ 2012 Sep;4(3):279-282 [FREE Full text] [doi: 10.4300/JGME-D-12-00156.1] [Medline: 23997866]
58. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
59. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. Br Med J 2010;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
60. Enders C. Applied Missing Data Analysis. New York, USA: Guilford Press; 2010.
61. Spring B, Schneider K, McFadden HG, Vaughn J, Kozak AT, Smith M, et al. Multiple behavior changes in diet and activity: a randomized controlled trial using mobile technology. Arch Intern Med 2012 May 28;172(10):789-796 [FREE Full text] [doi: 10.1001/archinternmed.2012.1044] [Medline: 22636824]
62. Prochaska JJ, Spring B, Nigg CR. Multiple health behavior change research: an introduction and overview. Prev Med 2008 Mar;46(3):181-188 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.001] [Medline: 18319098]
63. Thabane L, Mbuagbaw L, Zhang S, Samaan Z, Marcucci M, Ye C, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. BMC Med Res Methodol 2013 Jul 16;13(1):92 [FREE Full text] [doi: 10.1186/1471-2288-13-92] [Medline: 23855337]
64. Graham JW, Donaldson SI. Evaluating interventions with differential attrition: the importance of nonresponse mechanisms and use of follow-up data. J Appl Psychol 1993 Feb;78(1):119-128. [doi: 10.1037/0021-9010.78.1.119] [Medline: 8449850]
65. Blankers M, Smit ES, van der Pol P, de Vries H, Hoving C, van Laar M. The missing=smoking assumption: a fallacy in internet-based smoking cessation trials? Nicotine Tob Res 2016 Jan 5;18(1):25-33. [doi: 10.1093/ntr/ntv055] [Medline: 25744969]
66. Glasgow RE, Kurz D, King D, Dickman JM, Faber AJ, Halterman E, et al. Twelve-month outcomes of an Internet-based diabetes self-management support program. Patient Educ Couns 2012 Apr;87(1):81-92 [FREE Full text] [doi: 10.1016/j.pec.2011.07.024] [Medline: 21924576]
67. Soler RE, Leeks KD, Razi S, Hopkins DP, Griffith M, Aten A, Task Force on Community Preventive Services. A systematic review of selected interventions for worksite health promotion. The assessment of health risks with feedback. Am J Prev Med 2010 Feb;38(2 Suppl):S237-S262. [doi: 10.1016/j.amepre.2009.10.030] [Medline: 20117610]
68. Walthouwer MJL, Oenema A, Lechner L, de Vries H. Use and effectiveness of a video- and text-driven web-based computer-tailored intervention: randomized controlled trial. J Med Internet Res 2015 Sep 25;17(9):e222 [FREE Full text] [doi: 10.2196/jmir.4496] [Medline: 26408488]
69. Lorig K, Ritter PL, Laurent DD, Plant K, Green M, Jernigan VB, et al. Online diabetes self-management program: a randomized study. Diabetes Care 2010 Jun 18;33(6):1275-1281 [FREE Full text] [doi: 10.2337/dc09-2153] [Medline: 20299481]
70. McKay HG, King D, Eakin EG, Seeley JR, Glasgow RE. The diabetes network internet-based physical activity intervention: a randomized pilot study. Diabetes Care 2001 Aug 01;24(8):1328-1334. [doi: 10.2337/diacare.24.8.1328] [Medline: 11473065]

71. Hansel B, Giral P, Gambotti L, Lafourcade A, Peres G, Filipecki C, et al. A fully automated web-based program improves lifestyle habits and HbA1c in patients with type 2 diabetes and abdominal obesity: randomized trial of patient e-coaching nutritional support (the anode study). J Med Internet Res 2017 Nov 8;19(11):e360 [FREE Full text] [doi: 10.2196/jmir.7947] [Medline: 29117929]
72. Health Council of the Netherlands. Dutch physical activity guidelines 2017. The Hague: Health Council of the Netherlands; 2017. URL: https://www.rijksoverheid.nl/documenten/rapporten/2018/11/12/beweegrichtlijnen-2017 [accessed 2018-11-29]
73. Nicolaou M, Gademan MG, Snijder MB, Engelbert RH, Dijkshoorn H, Terwee CB, et al. Validation of the squash physical activity questionnaire in a multi-ethnic population: the Helius study. PLoS One 2016 Aug 30;11(8):e0161066 [FREE Full text] [doi: 10.1371/journal.pone.0161066] [Medline: 27575490]
74. DeWalt DA, Davis TC, Wallace AS, Seligman HK, Bryant-Shilliday B, Arnold CL, et al. Goal setting in diabetes self-management: taking the baby steps to success. Patient Educ Couns 2009 Nov;77(2):218-223 [FREE Full text] [doi: 10.1016/j.pec.2009.03.012] [Medline: 19359123]
75. Seligman HK, Wallace AS, DeWalt DA, Schillinger D, Arnold CL, Shilliday BB, et al. Facilitating behavior change with low-literacy patient education materials. Am J Health Behav 2007 Jul 1;31(1):69-78. [doi: 10.5993/ajhb.31.s1.9]
76. Schmeichel BJ, Vohs K. Self-affirmation and self-control: affirming core values counteracts ego depletion. J Pers Soc Psychol 2009 Apr;96(4):770-782. [doi: 10.1037/a0014635] [Medline: 19309201]
77. Finne E, Englert C, Jekauc D. On the importance of self-control strength for regular physical activity. Psychol Sport Exercise 2019 Jul;43:165-171. [doi: 10.1016/j.psychsport.2019.02.007]
78. van Osch L, Lechner L, Reubsaet A, De Vries H. From theory to practice: an explorative study into the instrumentality and specificity of implementation intentions. Psychol Health 2010 Mar;25(3):351-364. [doi: 10.1080/08870440802642155] [Medline: 20204939]
79. Nefs G, Bot M, Browne JL, Speight J, Pouwer F. Diabetes MILES--the Netherlands: rationale, design and sample characteristics of a national survey examining the psychosocial aspects of living with diabetes in Dutch adults. BMC Public Health 2012 Oct 30;12(1):925 [FREE Full text] [doi: 10.1186/1471-2458-12-925] [Medline: 23110382]
80. Malpass A, Andrews R, Turner KM. Patients with type 2 diabetes experiences of making multiple lifestyle changes: a qualitative study. Patient Educ Couns 2009 Feb;74(2):258-263. [doi: 10.1016/j.pec.2008.08.018] [Medline: 18848413]
81. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
82. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013 Feb 14;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
83. Van der Mispel C, Poppe L, Crombez G, Verloigne M, De Bourdeaudhuij I. A self-regulation-based ehealth intervention to promote a healthy lifestyle: investigating user and website characteristics related to attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]
84. Vermeire E, Hearnshaw H, Van Royen P, Denekens J. Patient adherence to treatment: three decades of research. A comprehensive review. J Clin Pharm Ther 2001 Oct;26(5):331-342. [doi: 10.1046/j.1365-2710.2001.00363.x] [Medline: 11679023]
85. Cunningham JA, Kypri K, McCambridge J. Exploratory randomized controlled trial evaluating the impact of a waiting list control design. BMC Med Res Methodol 2013 Dec 6;13(1):150 [FREE Full text] [doi: 10.1186/1471-2288-13-150] [Medline: 24314204]
86. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]

Abbreviations

FFQ: food frequency questionnaire
MDP: My Diabetes Profile
MET: metabolic equivalent of task
OHA: oral hypoglycemic agent
PA: physical activity
ProMAS: Probabilistic Medication Adherence Scale
RCT: randomized controlled trial
SQUASH: Short Questionnaire to Assess Health-Enhancing Physical Activity
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 03.03.20; peer-reviewed by AL Vuorinen, B Oh, A Psihogios, O Farris; comments to author 29.06.20; revised version received 17.09.20; accepted 21.12.20; published 23.02.21.

Please cite as:
Vluggen S, Candel M, Hoving C, Schaper NC, de Vries H
A Web-Based Computer-Tailored Program to Improve Treatment Adherence in Patients With Type 2 Diabetes: Randomized Controlled Trial
J Med Internet Res 2021;23(2):e18524
URL: https://www.jmir.org/2021/2/e18524
doi:10.2196/18524
PMID:33620321

Original Paper

Effects of Smartphone-Based Stress Management on Improving Work Engagement Among Nurses in Vietnam: Secondary Analysis of a Three-Arm Randomized Controlled Trial

Natsu Sasaki[1], MD; Kotaro Imamura[1], PhD; Thuy Thi Thu Tran[2], MSc; Huong Thanh Nguyen[3], PhD; Kazuto Kuribayashi[1]; Asuka Sakuraya[4], PhD; Thu Minh Bui[5], MNSc; Quynh Thuy Nguyen[2], PhD; Nga Thi Nguyen[3], MPH; Giang Thi Huong Nguyen[5], MNSc; Melvyn Weibin Zhang[6], MBBS; Harry Minas[7], FRANZCP; Yuki Sekiya[1], PhD; Kazuhiro Watanabe[1], PhD; Akizumi Tsutsumi[8], PhD; Akihito Shimazu[9], PhD; Norito Kawakami[1], MD, PhD

[1]Department of Mental Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[2]Department of Occupational Health and Safety, Faculty of Environmental and Occupational Health, Hanoi University of Public Health, Hanoi, Vietnam

[3]Faculty of Social Sciences—Behavior and Health Education, Hanoi University of Public Health, Hanoi, Vietnam

[4]Department of Public Health, Tokyo Women's Medical University, Tokyo, Japan

[5]Nursing Office, Bach Mai Hospital, Hanoi, Vietnam

[6]National Addiction Management Service, Institute of Mental Health, Singapore, Singapore

[7]Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Australia

[8]Department of Public Health, Kitasato University School of Medicine, Sagamihara, Japan

[9]Faculty of Policy Management, Keio University, Kanagawa, Japan

Corresponding Author:
Norito Kawakami, MD, PhD
Department of Mental Health
Graduate School of Medicine
The University of Tokyo
7-3-1, Hongo, Bunkyo-ku
Tokyo,
Japan
Phone: 81 3 5841 3522
Fax: 81 3 5841 3392
Email: nkawakami@m.u-tokyo.ac.jp

Abstract

Background: Work engagement is important for employee well-being and work performance. However, no intervention study has investigated the effect of an eMental Health intervention on work engagement among workers in low- and middle-income countries (LMICs).

Objective: The aim of the study was to examine the effects of a newly developed smartphone-based stress management program (ABC Stress Management) on improving work engagement among hospital nurses in Vietnam, an LMIC.

Methods: Full-time registered nurses (n=949) were randomly assigned to one of 2 intervention groups or a control group. The intervention groups were a 6-week, 6-lesson program offering basic cognitive behavioral therapy (CBT-based stress management skills), provided in either free-choice (program A) or fixed order (program B). Work engagement was assessed at baseline and 3-month and 7-month follow-ups in each of the 3 groups.

Results: The scores of work engagement in both intervention groups improved from baseline to 3-month follow-up, and then decreased at the 7-month follow-up, while the score steadily increased from baseline to 7-month follow-up in the control group. Program B showed a significant intervention effect on improving work engagement at the 3-month follow-up (P=.049) with a small effect size (Cohen d= 0.16; 95% CI 0.001 to 0.43]). Program A showed nonsignificant trend (d=0.13; 95% CI –0.014 to 0.41; P=.07) toward improved engagement at 3 months. Neither program achieved effectiveness at the 7-month follow-up.

Conclusions: The study demonstrated that a fixed order (program B) delivery of a smartphone-based stress management program was effective in improving work engagement in nurses in Vietnam. However, the effect was small and only temporary. Further

improvement of this program is required to achieve a greater effect size and more sustained, longer lasting impact on work engagement.

Trial Registration: University Hospital Medical Information Network Clinical Trials Registry UMIN000033139; tinyurl.com/55gxo253

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2018-025138

(J Med Internet Res 2021;23(2):e20445) doi:10.2196/20445

KEYWORDS

stress management; mental health; occupational health; digital health; workplace; LMICs; South-East Asia; health care professionals

Introduction

Work engagement is a popular topic in occupational mental health, given its positive impact on employee well-being and work performance [1,2]. Increasing and sustaining work engagement is therefore a prime concern of many organizations. Work engagement is one of the positive mental health outcomes resulting from a positive, fulfilling, work-related state of mind. It has 3 dimensions: vigor, dedication, and absorption [3]. Engaged workers who are connected to the work environment and its activities have sufficient energy to meet the demands of their job, perform well, and have low intention to leave [4]. Many individual and group-based [1,5] intervention programs have been developed to improve work engagement, including personal resource building (eg, resilience training), job resource building (eg, increased social support and feedback), leadership training (eg, skills improvement for managers), and health promotion (eg, stress management). A promising option would be an internet-delivered eHealth intervention that is feasible, low-cost, effective, and accessible [6-8]. An earlier randomized trial including an internet-based cognitive behavioral therapy (iCBT) intervention in the form of a self-care program showed a small effect size (d=0.16) and improved work engagement among workers [9]. Another web-based stress management literacy intervention that included CBT components also improved work engagement among participants with lower baselines of work engagement [10]. An iCBT intervention may be an affordable, sustainable, and effective approach to improving work engagement.

Smartphone ownership has been rapidly increasing in low- and middle-income countries (LMICs) [11], and they seem to be ready for eHealth implementation. In Vietnam in 2017, for instance, the percentage of people using smartphones in rural and urban areas was 68% and 84%, respectively [12]. In LMICs, work engagement has also been associated with better health, greater life satisfaction, and improved work performance [13-15]. An iCBT intervention may increase work engagement in low-resource settings such as LMICs. However, to date, no intervention study has investigated the effect of an iCBT or any other eMental Health intervention on work engagement among workers in LMICs.

Engagement in their work holds an important role in nurses' health, well-being [14,16], work performance; the quality of health care services they provide [17-19]; and turnover intentions [20-22]. In addition, nursing is a stressful profession with high work-related stress and burnout in both high-income countries and LMICs [23]. In Vietnam, job demands have increased because of the shortage of nurses [24-26] and the rapidly aging society [27-29]. Work engagement can help nurses manage work-related stress [17,30]. Two previous studies have investigated the effect of web-based interventions on improving nurses' work engagement in Germany, a high-income country [31,32]. The findings were inconsistent: one showed significant effectiveness [32] and the other did not [31]. It is important to know if an iCBT intervention will indeed increase work engagement among nurses in LMICs.

The purpose of this study was to determine whether a newly developed smartphone-based stress management program improved work engagement among hospital nurses in Vietnam. We analyzed data collected in a randomized controlled trial (RCT) that included work engagement as a secondary outcome [33].

Methods

Trial Design

This was a 3-armed RCT (allocation ratio: 1:1:1) examining improvement of depressive and anxiety symptoms as primary outcomes and, as stated, work engagement as a secondary outcome. The results of two internet-based CBT stress management programs (free-choice sequence and fixed sequential order) were compared with that of a control group at 3-month and 7-month follow-ups among hospital nurses in Vietnam [33]. The research ethics review board of the Graduate School of Medicine/Faculty of Medicine, University of Tokyo (No. 11991) and the ethical review board for Biomedical Research of Hanoi University of Public Health (No. 346/2018/YTCC-HD3) approved the study procedure. The study protocol was registered at the University Hospital Medical Information Network Clinical Trials Registry [UMIN000033139]. The trial protocol is published elsewhere [33]. This manuscript conforms to the Consolidated Standards of Reporting Trials (CONSORT) guidelines [34,35].

Participants

All participants (n=1256) were recruited from a large public tertiary hospital at national level in Hanoi, Vietnam in 2018. We distributed written information about the study, consent form, baseline questionnaires, and a numbered envelope to return the completed questionnaires anonymously. The inclusion criteria were being employed full time as a registered nurse and having internet access via a mobile device such as a smartphone. The exclusion criteria were (1) plans to change or leave the job

in the next 7 months, (2) being an assistant nurse and helper, (3) being temporary or part-time employed, (4) having taken leave for 15 or more days for a physical or mental condition in the past 3 months, and (5) undergoing treatment for a mental health problem from a mental health professional. However, exclusion criteria 4 and 5 were withdrawn before the start of the baseline survey (see Changes to the Protocol in Methods).

Intervention Programs

Two smartphone-based stress management programs (program A and program B) were developed in the ABC Stress Management app. Program A was a free-choice multimodule program, in which participants could select a module to complete each week in any order. Program B was a fixed-sequential order multimodule program, in which participants were required to complete one module per week in a fixed order. The contents of program A were based on a previous online stress management program to reduce the distress of office workers in Japan [36]. Program B included CBT-based stress management skills adapted from a previous iCBT program that reduced depressive symptoms in Japanese office workers [37]. Each program contained 6 modules. It took about 15 minutes to complete one module. We developed the programs based on discussions with Vietnamese nurses to consider the cultures and specific stressors that they could have at work. Several meetings were held to allow 30 head nurses to share their stressful experiences at work and their reflections on the draft program content; these head nurses were also invited to participate in reviewing the programs, and the programs were revised based on their feedback. Full details of these programs can be found in the published study protocol paper [33]. The mp4 file of the programs can be found in Multimedia Appendix 1.

Intervention Groups (Programs A and B)

Participants in the intervention groups were required to complete program A or B within 10 weeks after the baseline survey. Participants began the program after signing in with their ID and password. The clinical research coordinator sent weekly reminder messages to people who had not completed a module on time. An informal group chat (via social media apps such as Viber, Zalo, Facebook Messenger) with researchers and hospital head nurses was used to deliver intensive technical support for participants to complete the program. Before the start of the intervention, researchers helped participants download the app and complete an introductory module that provided a general explanation of how to use the app.

Control Group

Participants in the control group did not receive any intervention during the 7-month intervention period. However, they were free to use any other mental health service as usual treatment. Intervention programs were provided for the control group after the 7-month follow-up.

Outcome Measurement

All outcomes were assessed at baseline and, 3-month (the end of the intervention period), and 7-month follow-ups with a paper-based self-administered survey questionnaire. Participants were allowed 10 weeks to complete the program (6 weeks for the modules and 4 weeks to review any modules they desired). Follow-up at 3 months was timed to assess immediate effects of the interventions. For administrative reasons, we conducted a follow-up at 7 months, instead of a usual 6-month follow-up, to evaluate longer term effects.

Work Engagement

The short form of the Utrecht Work Engagement Scale–9 item (UWES-9) was used to assess work engagement [3]. The UWES-9 consists of 3 subscales (vigor, dedication, and absorption) that contain 3 items each. The UWES-9 uses a self-report 7-point rating scale (0= never; 6= every day). The mean scores of the 3 UWES subscales and the total score are computed by adding the scores and dividing the sum by the number of items in each subscale. Hence, the UWES's 3 subscale scores and a total score range from 0 to 6. The Vietnamese version of UWES-9 has been validated elsewhere [38]. The Cronbach alpha coefficients of the UWES-9 and vigor, absorption, and dedication subscales were .93, .86, .77, and .90, respectively. Confirmatory factor analyses indicated that the 3-factor structure was acceptable.

Demographic Variables

Demographic and occupational variables were assessed using a questionnaire and included gender (male or female), age, education (vocational school, college, university undergraduate, or postgraduate), marital status (single, married or divorced/widowed), and employment contract (fixed-time contract for less than one 1 year, fixed - time contract for more than 1 year, unlimited time contract, permanent contract, or other). Age was calculated based on the year of birth.

Sample Size Calculation

The sample size in the study was calculated for the primary outcome (ie, depressive symptoms assessed by the Depression, Anxiety, and Stress Scale–21-item [DASS-21]) when the effect size was set to 0.25. A post hoc test power (1-beta) for work engagement was calculated as 0.52 at the 3-month follow-up and 0.10 at the 7-month follow-up, assuming that the alpha was less than 0.05 (2-tailed) and the number of respondents in each group who were included in the analyses was 317, using the G*Power 3.1 program (Heinrich-Heine-Universität Düsseldorf) [39,40].

Randomization

Eligible participants were randomly assigned to one of the 3 trial arms (2 intervention groups and a control group). Stratified permuted-block randomization was conducted. The block sizes were fixed at 3. Participants were stratified according to the baseline depression subscale score of DASS-21 into 2 strata (≥10 or <10) [41]. A stratified permuted block random table was generated by an independent biostatistician. Enrollment was conducted by a clinical research coordinator (TTTT), and the assignment was conducted by an independent research assistant. The stratified permuted-block random table was password-protected and blinded to the researchers. Only the research assistant could access it during random allocation.

Statistical Analyses

For the main pooled analysis, a mixed model for repeated measures conditional growth model analysis with an unstructured covariance matrix was conducted using a group (intervention and control) × time (baseline and 3-month and 7-month follow-ups) interaction as an indicator of the intervention effect. The 2 intervention effects (program A vs control and program B vs control) were simultaneously tested in the model. For sensitivity analysis, a similar mixed model for repeated measures, but using the analysis of variance model with an unstructured covariance matrix, was conducted.

At the baseline survey, if the number of missing items was less than half of the number of total items, the missing values were imputed, using values calculated according to the following equation: (mean value × total number of items) / number of missing items. Otherwise, if the number of missing items was more than half of the total items, the case was not imputed and treated as missing. Missing values at follow-up surveys were imputed by applying the maximum likelihood estimation. An intention-to-treat principle was applied.

The effect size was estimated in 2 ways. First, we estimated a regression coefficient for a group (each of the 2 intervention groups vs the control group) × time (baseline and 2 follow-ups) interaction, which was converted to an effect size by dividing by a pooled standard deviation at baseline and follow-ups. Second, we calculated Cohen *d* among completers at baseline for each follow-up.

Statistical significance was defined as $P<.05$. All statistical analyses were performed using SPSS Statistics 26.0, Japanese version (IBM Corp).

Changes to the Protocol

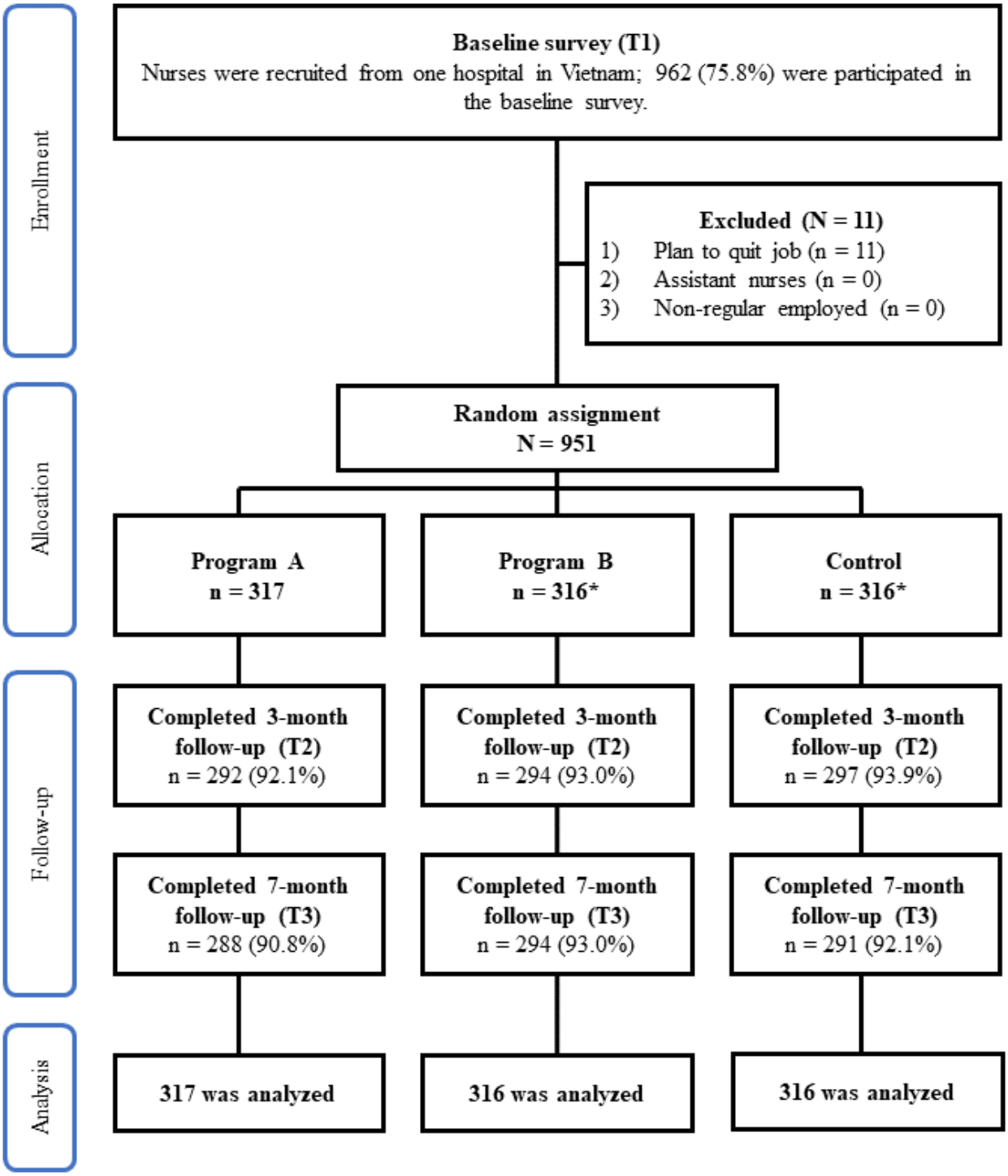
Before starting the study, 2 of the exclusion criteria (no. 4 and no. 5, described above) were removed because these restrictions would be expected to reduce the participation rate, and because reducing the eligibility restrictions when selecting participants is desirable for the purpose of ascertaining the effect of the programs in the real setting as a pragmatic trial [42].

Results

Characteristics of Participants

The participant flowchart is shown in Figure 1. In total, 75.81% (962/1269) of nurses participated in the baseline survey (September 2018). After 11 were excluded, 951 met the eligibility criteria. Finally, the 951 participants were randomly allocated with 317 in each group (2 interventions and one control group). After the random assignment, one participant in the intervention group (program B) and one in the control group were excluded because of duplicate registration. The follow-up rates were 92.11% to 93.04% (3 months, January 2019) and 90.85% to 93.04% (7 months, May 2019). The baseline characteristics of participants are shown in Table 1. Most (806/949, 84.9%) of the participants were female, nearly half (443/949, 46.7%) had graduated from vocational school, and the majority (793/949, 83.6%) were married. Employment contracts of most nurses were permanent (505/949, 53.2%). The average age was 33.1 years (range 22 to 58).

Figure 1. Participant flowchart.



* One was excluded after randomization because of duplicate registration.

Table 1. Demographic characteristics of study participants in the intervention and control groups of hospital nurses in Vietnam (n=949).

Characteristics	Program A (n=317)	Program B (n=316)	Control (n=316)
Gender, n (%)			
Male	39 (12.3)	56 (17.7)	48 (15.2)
Female	278 (87.7)	260 (82.3)	268 (84.4)
Age in years, mean (SD)	33.7 (7.3)	32.8 (6.6)	32.8 (6.4)
Education level, n (%)			
Vocational school	151 (47.6)	142 (44.9)	150 (47.5)
College	39 (12.3)	50 (15.8)	48 (15.2)
University undergraduate	118 (37.2)	119 (37.7)	111 (35.1)
Postgraduate	6 (1.9)	5 (1.6)	6 (1.9)
Unknown	3 (0.9)	0 (0)	1 (0.3)
Marital status, n (%)			
Single	40 (12.6)	44 (13.9)	53 (16.8)
Married	270 (85.2)	267 (84.5)	256 (81.0)
Divorced/widowed	5 (1.6)	5 (1.6)	6 (1.9)
Unknown	2 (0.6)	0 (0)	1 (0.3)
Employment contract, n (%)			
Fixed-time, ≤1 year	64 (20.2)	79 (25.0)	77 (24.4)
Fixed-time, >1 year	10 (3.2)	8 (2.5)	12 (3.8)
Unlimited time	62 (19.6)	69 (21.8)	63 (19.9)
Permanent	181 (57.1)	160 (50.6)	164 (32.5)

Effect of the Intervention Programs on Work Engagement

The average scores of work engagement and subscales at each time are shown in Table 2. The scores of work engagement in both intervention groups improved from baseline to 3-month follow-up but slightly decreased by the 7-month follow-up. The score in the control group increased at 3-month and 7-month follow-ups. Table 3 shows the estimated effect of both interventions on improving work engagement. Program B showed a significant estimated effect for improving work engagement at 3 months (t=1.97, P=.049, d=0.16; 95% CI 0.001 to 0.43). On the other hand, program A showed a nonsignificant trend (t=1.82, P=.069, d=0.13; 95% CI –0.014 to 0.41) toward improved work engagement at 3 months. At 7 months, neither program showed a significant effect (program A: t=0.02, P=.98, d=–0.004; 95% CI –0.17 to 0.16; program B: t=0.65, P=.52, d=0.05; 95% CI –0.11 to 0.21).

Table 2. The mean scores of the Utrecht Work Engagement Scale–9 item among intervention and control groups from baseline to 3-month and 7-month follow-ups and effect size (Cohen d).

Survey	Program A (n=317)		Program B (n=316)		Control (n=316)		Program A vs control	Program B vs control
UWES-9[a] (range 0-6)	n	mean (SD)	n	mean (SD)	n	mean (SD)	Cohen d[b] (95% CI)	Cohen d[b] (95% CI)
Baseline	317	4.3 (1.2)	316	4.3 (1.3)	316	4.3 (1.3)	—	—
3 months	292	4.6 (1.2)	294	4.7 (1.1)	297	4.4 (1.3)	0.13 (–0.03 to 0.29)	0.16 (0.002 to 0.32)[c]
7 months	288	4.5 (1.3)	294	4.6 (1.1)	290	4.5 (1.3)	–0.004 (–0.17 to 0.16)	0.05 (–0.11 to 0.21)

[a]UWES: Utrecht Work Engagement Scale–9 item.

[b]Effect sizes were calculated among completers at baseline for each follow-up.

[c]P=.049.

Table 3. The estimates of fixed effects and effect size at 3-month and 7-month follow-ups for work engagement in both intervention groups (programs A and B).

Time points	Estimates of fixed effects					Estimated effect size
	Effect size	SE	95% CI	*t* score	*P* value	
Program A						
3-month[a]	0.20	0.11	–0.02 to 0.41	1.82	.07	0.14
7-month[a]	0.003	0.12	–0.23 to 0.23	0.02	.98	0.002
Pooled[b]	0.005	0.06	–0.11 to 0.12	0.08	.94	—
Program B						
3-month[a]	0.21	0.11	0.001 to 0.43	1.97	.049	0.16
7-month[a]	0.08	0.12	–0.15 to 0.30	0.65	.52	0.05
Pooled[b]	0.04	0.06	–0.08 to 0.16	0.67	0.5	—

[a]Mixed-model for repeated measures analysis of variance model analyses was conducted.

[b]Mixed-model for repeated measures conditional growth model analyses was conducted.

Discussion

Principal Findings

The 3-arm RCT examined the effects of a newly developed smartphone-based stress management program to improve work engagement as the secondary outcome at 3-month and 7- month follow-ups among hospital nurses in Vietnam. Program B, a 6-module CBT program with fixed sequential order, showed a significant small intervention effect on work engagement at the 3-month follow-up (d=0.16, P=.049). Program A did not show a significant effect at the 3-month follow-up. Neither program achieved effectiveness at the 7-month follow-up. A smartphone-based stress management program with fixed sequential order may be effective in improving work engagement in a population of nurses in an LMIC, Vietnam.

Scores on the UWES-9 increased (indicating improved work engagement) from baseline to seven-month follow-up in the control group as well. A possible reason is potential contamination by information provided about the intervention programs from intervention groups to the control group [38]. Control group participants may have received information on stress management from their colleagues who were in the intervention groups. Other reasons may include a seasonal workload change or organizational reforms in the target hospital.

Program B significantly improved work engagement at the 3-month follow-up among nurses in the intervention groups compared with nurses in the control group. This result was consistent with a previous RCT on work engagement among information technology workers in Japan [9], which showed that a similar web-based, traditional CBT intervention with fixed-order modules achieved a significant but small intervention effect. This study expands the evidence on the effectiveness of iCBT on work engagement from workers in a high-income country [9] to hospital nurses in Vietnam, a LMIC.

In terms of the feasibility and reach of a smartphone-based intervention, more than 80% of nurses in the target hospital owned smartphones. The content of the programs was appropriate and relevant since the case story presented was based on interviews with nurses in Vietnam about managing their stress on the job. After making an outline, the content was reviewed and confirmed by nurses. These factors may facilitate the use of program B and support its ability to improve work engagement. However, the effectiveness of an iCBT program may depend on the culture and internet or smartphone literacy in a target country. Moreover, nurses are professionals. Their job requires an advanced college or university education. Additional research is needed to expand these findings to other LMICs and occupations.

Nurses who used program B were required to learn the cognitive behavioral model in an early module before continuing to later modules. Studying cognitive restructuring skills may decrease negative emotions [43], enhance positive cognitions concerning work-related challenges, and increase work engagement. It was reported that emotion-regulating cognitive styles, such as positive reinterpretation, focus on plans, acceptance, positive focus, and putting into perspective, are associated with work engagement [44]. Thus, cognitive styles changed by the cognitive restructuring skill may directly improve work engagement. It is also possible that adaptive coping skills learned from program B may increase personal resources (eg, self-efficacy [45]) and psychosocial job resources (eg, workplace social support) by helping employees achieve work goals and stimulate personal learning and development [46,47]. The mechanism by which an iCBT program improves work engagement needs to be clarified in future research.

There was no significant effect of program B on work engagement at 7 months. A previous review of eHealth interventions indicated that the long-term benefits of an eHealth program are still unknown due to small effect size [48]. In this study, it is plausible that participants forgot what they had learned because of its lack of intensity and repeatability. To improve long-term effectiveness, refresher sessions may be needed after the intervention period. Another reason might be

the small sample size for proving the effectiveness on work engagement. A post hoc power analysis revealed that a sample of 615 in a group might be needed to detect the effect size (d=0.16) for 1-beta over .80. Further study should be conducted to examine long-term effectiveness for hospital nurses.

Program A shows only nonsignificant effect on work engagement at 3-month (d=0.16, P=.07) and 7-month follow-up. One study in a Western country reported that tailored (ie, partially personalized) options provided a better outcome of eHealth programs because free choice was culturally preferred [49]. Our result was inconsistent with that study. Perhaps the more collectivist and hierarchically oriented cultures in Asian countries are associated with people feeling more comfortable doing as others do and following instructions [50]. The free-choice style of program A might be a psychological burden for participants in Vietnam since they had to think about which module to read each week. Another reason was the difference in content between the 2 programs: program A did not intensively repeat cognitive restructuring skills or follow a sequential learning order of standardized CBT due to free-choice operation. A process evaluation showed similar completion rates for both programs (83.3% for program A and 86.1% for program B). A significant difference of process evaluation such as satisfaction and usefulness was not found (available upon request). The free-choice order program without intensive restructuring skills might not have a significant effect on Vietnamese nurses.

Strengths and Limitations

The strengths of this RCT are the type of program delivery and high completion rate. The iCBT programs used in this study is fully automated and self-guided. Considering the small effect size, a self-guided program still has the merits of greater accessibility and lower cost than interactive or specialist-guided programs. This study had a high completion rate. For evaluating the cost-effectiveness and sustainability of such a program, it will be necessary to observe completion rates in practical settings without intensive reminders. Despite the small effect size, program B appears to be useful in improving nurses' work engagement, with expected benefits for their health, quality of their services, and patient clinical outcomes. Successful dissemination and implementation of internet-delivered stress management for nurses will require collaboration between policymakers and stakeholders.

The study has several limitations. First, participants were recruited from a large and prestigious national general hospital in Hanoi, and were limited to full-time nurses with a personal smartphone. Therefore, the generalizability of these findings to the wider nursing population is limited. Second, this study did not adopt a stratified randomized method by using work engagement scores. Therefore, we did not conduct a subgroup analysis. Further studies on improving work engagement by stratified randomization might achieve a larger effect size in the low work engagement population. Third, all outcomes in this study were measured by self-report, which might be affected by participant perceptions or institutional factors. Future studies should consider the use of additional objective outcome measures such as supervisor ratings of work performance. Fourth, the possibility of contamination of information to the control group may have reduced differences between intervention and control groups, resulting in possible underestimation of intervention effectiveness not fully controlled in this study. Fifth, besides the intervention programs, an informal group chat (via Viber, Zalo, Facebook Messenger) led by researchers and hospital head nurses used to increase the participation rate may also have contributed to improvement of work engagement in the intervention groups. Sixth, there may have been social pressure or frustrations caused by frequent reminding to study the programs, reducing the effect of the program on work engagement. Seventh, the programs were designed mainly to target smartphone users, but they can be used on a PC or tablet. Nurses who did not have smartphones or internet access were excluded from access to the programs. In a future trial, in addition to the smartphone-based program, the same content in the programs should also be provided via other delivery modes such as computer, tablet, or booklet. Finally, it is not known whether a similar effect of this program would be observed outside of the nursing profession, in Vietnam or in other LMICs. Future studies should explore the generalizability of our findings to occupations other than nurses and in other LMICs.

Conclusion

The study demonstrated that a smartphone-based stress management program with fixed order significantly improved work engagement in a working population of nurses in an LMIC, Vietnam. However, the effect of the intervention was small and temporary. Further improvement of this program is necessary to achieve a greater effect size and more sustained impact on work engagement. However, fully automated and self-guided programs with great accessibility and minimal cost can be well suited for LMICs. The generalizability of these findings to occupations other than nursing and in other LMICs should be investigated in future studies.

Acknowledgments

This research was supported by the Japan Agency for Medical Research and Development under grant numbers JP17jk0110014 and JP18jk0110014. The funder had no role in study design, data collection, data analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

NS, KI, TTTT, HTN, KK, AS, TMB, AQN, QTN, KTN, GTHN, XTNT, TQT, MWBZ, HM, YS, AT, and NK conceived and designed the experiments. NS, KI, TTTT, HTN, KK, AS, YS, and NK contributed reagents, materials, and analytical tools. TTT

and HTN conducted the intervention in local settings. NS, KI, TTTT, HTN, HM, and NK wrote the paper. All authors read and approved the final paper.

Conflicts of Interest

NK reports grants from Infocom Corp, Fujitsu Ltd, Fujitsu Software Technologies, and TAK Ltd and personal fees from the Occupational Health Foundation, Japan Dental Association, Sekisui Chemicals, Junpukai Health Care Center, and Osaka Chamber of Commerce and Industry outside the submitted work.

Multimedia Appendix 1
The ABC stress management program.
[MP4 File (MP4 Video), 15973 KB - jmir_v23i2e20445_app1.mp4]

Multimedia Appendix 2
CONSORT-EHEALTH V1.6.1 checklist.
[PDF File (Adobe PDF File), 96 KB - jmir_v23i2e20445_app2.pdf]

References

1. Christian M, Garza A, Slaughter J. Work engagement: a quantitative review and test of its relations with task and contextual performance. Personnel Psychol 2011;64(1):89-136. [doi: 10.1111/j.1744-6570.2010.01203.x]
2. Halbesleben J. A meta-analysis of work engagement: relationships with burnout, demands, resources, and consequences. In: Work Engagement: A Handbook of Essential Theory and Research. London: Routledge; 2010:102-117.
3. Schaufeli WB, Salanova M, González-romá V, Bakker AB. The measurement of engagement and burnout: a two sample confirmatory factor analytic approach. J Happiness Stud 2002;3(1):71-92. [doi: 10.1023/A:1015630930326]
4. Halbesleben JR, Wheeler AR. The relative roles of engagement and embeddedness in predicting job performance and intention to leave. Work Stress 2008 Jul;22(3):242-256. [doi: 10.1080/02678370802383962]
5. Knight C, Patterson M, Dawson J. Work engagement interventions can be effective: a systematic review. Eur J Work Org Psychol 2019 Apr 06;28(3):348-372. [doi: 10.1080/1359432x.2019.1588887]
6. Carolan S, Harris PR, Cavanagh K. Improving employee well-being and effectiveness: systematic review and meta-analysis of web-based psychological interventions delivered in the workplace. J Med Internet Res 2017 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
7. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. PLoS One 2017;12(12):e0189904 [FREE Full text] [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]
8. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]
9. Imamura K, Kawakami N, Furukawa TA, Matsuyama Y, Shimazu A, Umanodan R, et al. Effects of an internet-based cognitive behavioral therapy intervention on improving work engagement and other work-related outcomes: an analysis of secondary outcomes of a randomized controlled trial. J Occup Environ Med 2015 May;57(5):578-584. [doi: 10.1097/JOM.0000000000000411] [Medline: 25749132]
10. Imamura K, Kawakami N, Tsuno K, Tsuchiya M, Shimada K, Namba K, et al. Effects of web-based stress and depression literacy intervention on improving work engagement among workers with low work engagement: an analysis of secondary outcome of a randomized controlled trial. J Occup Health 2017 Jan 24;59(1):46-54 [FREE Full text] [doi: 10.1539/joh.16-0187-OA] [Medline: 27885247]
11. Silver L. Chapter 2: In emerging economies, smartphone adoption has grown more quickly among younger generations. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. Washington: Pew Internet and American Life Project; 2019 Feb 05. URL: https://www.pewresearch.org/global/wp-content/uploads/sites/2/2019/02/Pew-Research-Center_Global-Technology-Use-2018_2019-02-05.pdf [accessed 2020-03-10]
12. Nielsen Vietnam smartphone insight report Q4 2017. URL: https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/Web_Nielsen_Smartphones20Insights_EN.pdf [accessed 2020-01-20]
13. Demerouti E, Mostert K, Bakker AB. Burnout and work engagement: a thorough investigation of the independency of both constructs. J Occup Health Psychol 2010 Jul;15(3):209-222. [doi: 10.1037/a0019408] [Medline: 20604629]
14. Panthee B, Shimazu A, Kawakami N. Validation of Nepalese version of Utrecht Work Engagement Scale. J Occup Health 2014;56(6):421-429 [FREE Full text] [doi: 10.1539/joh.14-0041-OA] [Medline: 25214188]
15. Sohrabizadeh S, Sayfouri N. Antecedents and consequences of work engagement among nurses. Iran Red Crescent Med J 2014 Nov;16(11):e16351 [FREE Full text] [doi: 10.5812/ircmj.16351] [Medline: 25763212]
16. Fong TC, Ng S. Measuring engagement at work: validation of the Chinese version of the Utrecht Work Engagement Scale. Int J Behav Med 2012 Sep;19(3):391-397 [FREE Full text] [doi: 10.1007/s12529-011-9173-6] [Medline: 21681564]

17. Mauno S, Ruokolainen M, Kinnunen U, De Bloom J. Emotional labour and work engagement among nurses: examining perceived compassion, leadership and work ethic as stress buffers. J Adv Nurs 2016 May;72(5):1169-1181. [doi: 10.1111/jan.12906] [Medline: 26841277]
18. Van Bogaert P, Peremans L, Van Heusden D, Verspuy M, Kureckova V, Van de Cruys Z, et al. Predictors of burnout, work engagement and nurse reported job outcomes and quality of care: a mixed method study. BMC Nurs 2017;16:5 [FREE Full text] [doi: 10.1186/s12912-016-0200-4] [Medline: 28115912]
19. Han S, Han J, Kim Y. Effect of nurses' emotional labor on customer orientation and service delivery: the mediating effects of work engagement and burnout. Saf Health Work 2018 Dec;9(4):441-446 [FREE Full text] [doi: 10.1016/j.shaw.2017.12.001] [Medline: 30559993]
20. De Simone S, Planta A. The intention to leave among nurses: the role of job satisfaction, self-efficacy and work engagement. Med Lav 2017 Apr 21;108(2):87-97. [doi: 10.23749/mdl.v108i2.6074] [Medline: 28446735]
21. Marques-Pinto A, Jesus EH, Mendes AMOC, Fronteira I, Roberto MS. Nurses' intention to leave the organization: a mediation study of professional burnout and engagement. Span J Psychol 2018 Aug 08;21:E32. [doi: 10.1017/sjp.2018.30] [Medline: 30088463]
22. Moloney W, Boxall P, Parsons M, Cheung G. Factors predicting registered nurses' intentions to leave their organization and profession: a job demands-resources framework. J Adv Nurs 2018 Apr;74(4):864-875. [doi: 10.1111/jan.13497] [Medline: 29117451]
23. Salmasi S, Khan TM, Hong YH, Ming LC, Wong TW. Medication errors in the southeast asian countries: a systematic review. PLoS One 2015;10(9):e0136545 [FREE Full text] [doi: 10.1371/journal.pone.0136545] [Medline: 26340679]
24. Kanchanachitra C, Lindelow M, Johnston T, Hanvoravongchai P, Lorenzo FM, Huong NL, et al. Human resources for health in southeast Asia: shortages, distributional challenges, and international trade in health services. Lancet 2011 Feb 26;377(9767):769-781. [doi: 10.1016/S0140-6736(10)62035-1] [Medline: 21269674]
25. Elison NK, Verani AR, McCarthy C. National nursing and midwifery legislation in countries of South-East Asia with high HIV burdens. WHO South East Asia J Public Health 2015 Jan;4(102):12-19 [FREE Full text] [doi: 10.4103/2224-3151.206615] [Medline: 26568920]
26. The world health report 2006: working together for health. World Health Organization. URL: https://www.who.int/whr/2006/whr06_en.pdf?ua=1 [accessed 2021-01-15]
27. Toh SG, Ang E, Devi MK. Systematic review on the relationship between the nursing shortage and job satisfaction, stress and burnout levels among nurses in oncology/haematology settings. Int J Evid Based Healthc 2012 Jun;10(2):126-141. [doi: 10.1111/j.1744-1609.2012.00271.x] [Medline: 22672602]
28. Fleming KC, Evans JM, Chutka DS. Caregiver and clinician shortages in an aging nation. Mayo Clin Proc 2003 Aug;78(8):1026-1040. [doi: 10.4065/78.8.1026] [Medline: 12911051]
29. Duvall JJ, Andrews DR. Using a structured review of the literature to identify key factors associated with the current nursing shortage. J Prof Nurs 2010;26(5):309-317. [doi: 10.1016/j.profnurs.2010.02.002] [Medline: 20869031]
30. Pérez-Fuentes MDC, Molero Jurado MDM, Barragán Martín AB, Simón Márquez MDM, Martos Martínez A, Gázquez Linares JJ. The mediating role of perceived stress in the relationship of self-efficacy and work engagement in nurses. J Clin Med 2018 Dec 21;8(1):10 [FREE Full text] [doi: 10.3390/jcm8010010] [Medline: 30577596]
31. Kloos N, Drossaert CHC, Bohlmeijer ET, Westerhof GJ. Online positive psychology intervention for nursing home staff: a cluster-randomized controlled feasibility trial of effectiveness and acceptability. Int J Nurs Stud 2019 Oct;98:48-56. [doi: 10.1016/j.ijnurstu.2019.06.004] [Medline: 31295708]
32. Gollwitzer PM, Mayer D, Frick C, Oettingen G. Promoting the self-regulation of stress in health care providers: an internet-based intervention. Front Psychol 2018;9:838 [FREE Full text] [doi: 10.3389/fpsyg.2018.00838] [Medline: 29962979]
33. Imamura K, Tran TTT, Nguyen HT, Kuribayashi K, Sakuraya A, Nguyen AQ, et al. Effects of two types of smartphone-based stress management programmes on depressive and anxiety symptoms among hospital nurses in Vietnam: a protocol for three-arm randomised controlled trial. BMJ Open 2019 Apr 08;9(4):e025138 [FREE Full text] [doi: 10.1136/bmjopen-2018-025138] [Medline: 30962230]
34. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS Consensus Group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2:64 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
35. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
36. Sakuraya A, Imamura K, Kagami H. Which e-learning stress management program improve psychological distress most? A multi-arm randomized controlled trial. 2017 Presented at: The 90th Annual Meeting of Japan Society for Occupational Health; 2017; Tokyo.
37. Imamura K, Kawakami N, Furukawa TA, Matsuyama Y, Shimazu A, Umanodan R, et al. Effects of an Internet-based cognitive behavioral therapy (iCBT) program in Manga format on improving subthreshold depressive symptoms among healthy workers: a randomized controlled trial. PLoS One 2014;9(5):e97167 [FREE Full text] [doi: 10.1371/journal.pone.0097167] [Medline: 24844530]

38. Tran TTT, Watanabe K, Imamura K, Nguyen HT, Sasaki N, Kuribayashi K, et al. Reliability and validity of the Vietnamese version of the 9‐item Utrecht Work Engagement Scale. J Occup Health 2020 Aug 22;62(1):e12157. [doi: 10.1002/1348-9585.12157]
39. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
40. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
41. Tran TD, Tran T, Fisher J. Validation of the Depression Anxiety Stress Scales (DASS) 21 as a screening instrument for depression and anxiety in a rural community-based cohort of northern Vietnamese women. BMC Psychiatry 2013 Jan 12;13:24 [FREE Full text] [doi: 10.1186/1471-244X-13-24] [Medline: 23311374]
42. Ford I, Norrie J. Pragmatic trials. N Engl J Med 2016 Aug 04;375(5):454-463. [doi: 10.1056/NEJMra1510059] [Medline: 27518663]
43. Beck A. Cognitive Therapy of Depression. New York: Guilford Press; 1979.
44. Castellano E, Muñoz-Navarro R, Toledo MS, Spontón C, Medrano LA. Cognitive processes of emotional regulation, burnout and work engagement. Psicothema 2019 Feb;31(1):73-80. [doi: 10.7334/psicothema2018.228] [Medline: 30664414]
45. Lorente L, Salanova M, Martínez IM, Vera M. How personal resources predict work engagement and self-rated performance among construction workers: a social cognitive perspective. Int J Psychol 2014 Jun;49(3):200-207. [doi: 10.1002/ijop.12049] [Medline: 24821509]
46. Knight C, Patterson M, Dawson J. Building work engagement: a systematic review and meta-analysis investigating the effectiveness of work engagement interventions. J Organ Behav 2017 Jul;38(6):792-812 [FREE Full text] [doi: 10.1002/job.2167] [Medline: 28781428]
47. Bakker AB, Demerouti E. Towards a model of work engagement. Career Dev Int 2008 May 09;13(3):209-223. [doi: 10.1108/13620430810870476]
48. Ruwaard J, Lange A, Schrieken B, Dolan CV, Emmelkamp P. The effectiveness of online cognitive behavioral treatment in routine clinical practice. PLoS One 2012;7(7):e40089 [FREE Full text] [doi: 10.1371/journal.pone.0040089] [Medline: 22792217]
49. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J Med Internet Res 2009;11(2):e15 [FREE Full text] [doi: 10.2196/jmir.1151] [Medline: 19632969]
50. Markus HR, Kitayama S. Culture and the self: implications for cognition, emotion, and motivation. Psychol Rev 1991 Apr;98(2):224-253. [doi: 10.1037/0033-295X.98.2.224]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
DASS-21: Depression, Anxiety, and Stress Scale–21-item
iCBT: internet-based cognitive behavioral therapy
LMICs: low- and middle-income countries
RCT: randomized controlled trial
UMIN: University Hospital Medical Information Network Clinical Trials Registry
UWES-9: Utrecht Work Engagement Scale–9 item

Edited by T Rashid Soron; submitted 19.05.20; peer-reviewed by G Kian Liang, T Ingviya, C Miranda; comments to author 20.06.20; revised version received 03.09.20; accepted 17.10.20; published 23.02.21.

Please cite as:
Sasaki N, Imamura K, Tran TTT, Nguyen HT, Kuribayashi K, Sakuraya A, Bui TM, Nguyen QT, Nguyen NT, Nguyen GTH, Zhang MW, Minas H, Sekiya Y, Watanabe K, Tsutsumi A, Shimazu A, Kawakami N
Effects of Smartphone-Based Stress Management on Improving Work Engagement Among Nurses in Vietnam: Secondary Analysis of a Three-Arm Randomized Controlled Trial
J Med Internet Res 2021;23(2):e20445
URL: https://www.jmir.org/2021/2/e20445
doi:10.2196/20445
PMID:33620328

Original Paper

The Promise and Peril of Mobile Phones for Youth in Rural Uganda: Multimethod Study of Implications for Health and HIV

Philip Kreniske[1,2], PhD; Alyssa Basmajian[3], MA; Neema Nakyanjo[4], MA; William Ddaaki[4], BA, MSc; Dauda Isabirye[4], BA; Charles Ssekyewa[4], BA; Rosette Nakubulwa[4], BA; Jennifer S Hirsch[3], BA, PhD; Andrea Deisher[2], BSN; Fred Nalugoda[4], PhD; Larry W Chang[4,5,6], MD, MPH; John S Santelli[2,7], MD, MPH

[1]HIV Center for Clinical and Behavioral Studies, New York State Psychiatric Institute and Columbia University, New York, NY, United States

[2]Heilbrunn Department of Population and Family Health, Mailman School of Public Health, Columbia University, New York, NY, United States

[3]Sociomedical Sciences, Mailman School of Public Health, Columbia University, New York, NY, United States

[4]Rakai Health Sciences Program, Kalisizo, Uganda

[5]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[6]Division of Infectious Diseases, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, United States

[7]Pediatrics, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, United States

Corresponding Author:
Philip Kreniske, PhD
HIV Center for Clinical and Behavioral Studies
New York State Psychiatric Institute and Columbia University
722 West 168th Street
New York, NY, New York
United States
Phone: 1 646 774 6947
Fax: 1 646 774 6955
Email: pk2361@columbia.edu

Abstract

Background: In East Africa, where landlines are used by 1% of the population and access to the internet is limited, owning a cell phone is rapidly becoming essential for acquiring information and resources. Our analysis illuminates the perils and potential promise of mobile phones with implications for future interventions to promote the health of adolescents and young adults (AYAs) and to prevent HIV infection.

Objective: The aim of this study is to describe the current state of AYAs' phone use in the region and trace out the implications for mobile health interventions.

Methods: We identified 2 trading centers that were representative of southern Uganda in terms of key demographics, proportion of cell phone ownership, and community HIV prevalence. We stratified the sample of potential informants by age group (15-19 years and 20-24 years), gender, and phone ownership and randomly sampled 31 key informant interview participants within these categories. In addition, we conducted 24 ethnographic participant observations among AYAs in the communities of study.

Results: AYA frequently reported barriers to using their phones, such as difficulty accessing electricity. Nearly all AYAs used mobile phones to participate in the local economy and communicate with sexual partners. Phone use was frequently a point of contention between sexual partners, with many AYAs reporting that their sexual partners associated phone use with infidelity. Few AYAs reported using their phones for health-related purposes, with most getting health information in person from health workers. However, most AYAs reported an instance when they used their phone in an emergency, with childbirth-related emergencies being the most common. Finally, most AYAs reported that they would like to use their phones for health purposes and specifically stated that they would like to use their mobile phones to access current HIV prevention information.

Conclusions: This study demonstrates how mobile phones are related to income-generating practices in the region and communication with sexual partners but not access to health and HIV information. Our analysis offers some explanation for our previous study, which suggested an association between mobile phone ownership, having multiple sexual partners, and HIV risk. Mobile phones have untapped potential to serve as tools for health promotion and HIV prevention.

(J Med Internet Res 2021;23(2):e17837) doi:10.2196/17837

KEYWORDS

adolescence; youth; mobile phones; cell phones; mobility; HIV; East Africa

Introduction

Background

In East Africa, where landlines are used by 1% of the population [1] and access to the internet is limited [2], owning a mobile phone is rapidly becoming essential for acquiring information and resources [1]. In the past 10 years, mobile phone ownership in East Africa doubled from 30% to over 70% [2], with recent analyses indicating similar trends in the Ugandan region of study [3]. Thus, mobile phones may present an opportunity for contacting and providing health information to adolescents and young adults (AYAs) and referring them to health care services [4-6]. However, ethnographic [7,8] and quantitative research in East Africa have also documented the association between mobile phones and sexual behaviors [3]. This has implications for health as East Africa is among the regions with the greatest number of new HIV infections in the world, with AYAs at greatest risk [9]. The objective of this study is to describe the state of AYAs' phone use in a rural region of southern Uganda and identify the implications for HIV prevention and mobile health interventions.

Perils of Mobile Phones

In Uganda, our recent cross-sectional analysis of the Rakai Community Cohort Study (RCCS) suggested that after adjusting for demographic characteristics, including socioeconomic status, people who owned mobile phones were more likely to have multiple partners than people who did not have mobile phones and that young women (15-24 years) who owned mobile phones were less likely to use condoms consistently, more likely to consume alcohol before sex, and more likely to be HIV positive [3]. However, the mechanism for this increased risk was not explained, and we called for further qualitative examination of the issue.

Previous ethnographic research has investigated the proliferation of mobile phones in relation to sexual intimacy and HIV risk. One ethnographic study found that university students in Tanzania used cell phones to maintain privacy in romantic relationships and to engage in transactional sex, defined as the exchange of sex for material resources or symbolic capital [10]. Ethnographic research with boarding school students in Kenya found that men gifted phones to female classmates expressly to schedule future meetings and engage in sexual relations outside of school [11]. These studies highlight the need to better understand the mechanisms connecting mobile phone ownership and sexual risk.

Promise of Mobile Phones

The majority of mobile phone–based health interventions in low-to-middle-income countries (LMICs) leverage SMS text messaging [12,13]. A review found mobile phone–based interventions for adolescent sexual and reproductive health (SRH) in LMIC to be an effective way to reach young people, increase health knowledge, and achieve behavior change and noted a need for more research on security, confidentiality, and structural issues such as phone access [12]. More generally, several systematic reviews have shown that there is substantial evidence for the utility of mobile phone reminders to improve antiretroviral therapy adherence for people living with HIV in East Africa [14,15]. Studies in Uganda suggest that SMS text messaging is also a feasible and acceptable platform to support tuberculosis medication adherence [16,17] and testing [18,19]. Given the rapid rise in mobile phone ownership and the types of phones available and the importance of tailored interventions, there is a pressing need to examine the current state of mobile phones among AYAs in East Africa.

Methods

Data Sources

Background of the Research Setting

The greater Rakai region in southern Uganda is largely rural with dispersed trading centers. This paper focuses on the trading centers that have higher concentrations of AYAs and higher HIV prevalence than agricultural communities [20]. In both trading centers and agricultural communities, agricultural work is the main source of income and subsistence for most of the population [3]. The region borders Lake Victoria in the east and Tanzania in the south. The population is approximately 516,000, most people (57%) are aged between 0 and 17 years, with 19% aged between 18 and 30 years and 24.2% aged 31 years and older [21]. HIV prevalence from 2011 to 2013 ranged from 10% to 25%, depending on the community [20], and adjusted HIV incidence rate was 0.58 cases per 100 person years [9].

National data indicate that 73% of households have a mobile phone [22], which is comparable with the proportion of people in the rural region of the study who own phones [3]. However, only 16% of mobile phone users own smartphones [23]. In 2016-2017, the national unemployment rate was 9.7% and the median monthly income for employed residents in rural Uganda was UGX 150,000 (US $39.88) [24]. Ugandans who own mobile phones spend an average of UGX 14,500 (US $3.85) per month on voice calls [25], which is nearly 10% of the median monthly earnings for a rural resident. For smartphone users, 1 gigabyte of prepaid data generally costs UGX 10,265 (US $2.77) [23]. In addition, in 2018, the Ugandan parliament passed into legislation the over-the-top tax (OTT) of UGX 185 (US $0.05) per day to use social media services [26]. Although the cost of maintaining cell phone ownership in relation to median monthly earnings varies greatly based on region, rural residents generally endure a higher cost of cell phone ownership in relation to their income. Many mobile phone users use pay-as-you-go services [27], buying SIM cards, often from multiple networks [28], to pay for calls and SMS text messaging services when disposable income is available.

RCCS Quantitative Data

We used RCCS demographic data from 2018 to identify key informant interview (KII) participants. Beginning in 1994 and continuing to date, the RCCS records demographics and SRH

responses from an open cohort of residents aged 15-49 years from 40 communities in the greater Masaka region of Uganda. The RCCS research design and procedures have been described in detail elsewhere [9,29].

In this study, the RCCS data set enabled our team to purposively sample participants for KIIs. On the basis of a series of meetings with local experts, field observations, and a review of RCCS data, we identified 2 trading centers that were representative of the region in terms of key demographics, including the proportion of AYAs to the total trading center population, distribution of occupations, proportion of cell phone ownership, and community HIV prevalence. We then stratified the sample of potential KII participants by age group (15-19 years and 20-24 years), gender, and phone ownership and randomly sampled KII participants within these demographic categories.

The Rakai Health Sciences Program's Social and Behavioral Sciences Team

Social and Behavioral Sciences (SBS) team members began qualitative data collection in June 2018 and gathered the majority of data between January and May 2019. The SBS team members (authors 3-7) were fluent in English and Luganda (the most widely used language in the region). Most team members lived in the region of study and had 5 to 10 years of experience conducting qualitative research.

Semistructured KII

Drawing on narrative theory [30-32], the research team collaboratively designed and revised a series of KII questions that would elicit stories about participants' lives. The narrative approach posits that people communicate their beliefs and practices through stories and that systematic analysis of stories can illuminate structural and social dynamics. For example, in this study, asking participants to tell a story about a time they could not use their phone might illuminate structural barriers, such as inconsistent network coverage, or social dynamics such as which members of society were permitted to own and use phones. Thus, in this study, the KII questions were designed to prompt AYAs to tell stories about their mobile phone practices as they related broadly to 4 main domains of interest: socializing practices, mobile phone use and barriers, behaviors with intimate partners, and health information–seeking behaviors. Multimedia Appendix 1 contains the full list of interview questions and probes.

Ethnographic Participant Observations

Ethnographic participant observation (EPO) involves engaging with and recording people's daily lives [33]. Before conducting EPO, the SBS team created social maps of the communities of study with a focus on AYAs [34,35]. On the basis of these social maps and subsequent discussions, we identified 7 locations in which to conduct EPO: 1 hair salon, 1 boda (inexpensive and commonly used motorcycle taxi) stage (a place where boda boda drivers congregate and wait for customers), 2 bars, and 2 outdoor pool tables. At each location, the researchers identified an index case who served as the main contact and was present at each visit. SBS team members selected index cases who they determined to be regulars at the study site. SBS researchers visited each location 3 to 4 times over a 1-month period, spending approximately 2 hours per visit and observing and interacting with participants who they documented using field notes [36].

Analysis

This study analyzes KII and EPO data using grounded theory to present detailed characterizations of how AYAs in peri-urban locations were using mobile phones with implications for health risks and future interventions [33,37]. The iterative process took the form of a data analysis spiral [38] such that following data collection, we organized the data, read and memoed emerging ideas, described and classified codes, developed and assessed interpretations, and finally created an account of our analysis in the form of this research report.

Authors 1 and 2 read all the KII and EPO field notes several times and compiled the data in the qualitative software program Dedoose [39]. Using Dedoose, author 1 wrote and discussed the initial memos with author 2 and then created preliminary codes. At this point, as there were increasing cases of the same codes, with few new codes emerging, author 1 and author 2 believed thematic saturation had been achieved [37,40]. Author 1 then discussed all preliminary codes with all authors in a series of meetings in New York and Uganda, and SBS team members (authors 3-7) provided additional memos and codes. Authors 1 and 2 refined the preliminary codes, and using 20% of the data, they achieved 90% interrater reliability.

Ethical Considerations

Approvals for this study were obtained from the Research and Ethics Committee of the Uganda Virus Research Institute and the Uganda National Council for Science and Technology and from the Institutional Review Boards at Columbia University.

Results

Organizational Note on Results

In each section, we first present the general patterns identified in the KII and support these with key excerpts from the KII and, in some cases, EPO vignettes.

Participants

We included both AYAs who owned at least one mobile phone (n=22) and AYA who did not own a mobile phone but did have access to a shared phone (n=8) in the KII (Table 1). Our intention in interviewing AYAs who did not own phones was to compare their beliefs and reported behaviors with those of AYAs who did own phones.

Table 1. Characteristics of 15- to 24-years-olds who participated in key informant interviews.

Characteristics	Values, n (%)	
	Men	Women
Number of interviews	14 (45)	17 (55)
Mobile phone status		
Owns a mobile phone	10 (71)	13 (77)
Does not own a mobile phone	4 (29)	4 (24)
Age (years)		
15-19	7 (50)	6 (35)
20-24	7 (50)	11 (65)
Primary occupation		
Housework	0 (0)	4 (24)
Unemployed	0 (0)	1 (6)
Student	8 (57)	8 (47)
Business	0 (0)	3 (18)
Other (driver, carpenter, etc)	6 (43)	1 (6)

Phone Use Practices

In our final KII sample, most of the participants owned at least one phone (n=17), with 6 participants owning 2 phones and 6 participants only having access to but not owning a phone, a common practice in the region [41]. In the cases where participants owned 2 phones, 1 was a *flip phone* or often referred to in the region as a *button phone* and 1 was a smartphone. A button phone can be used to make calls and send text messages and occasionally has a very basic camera, and according to AYAs, some button phone batteries last over a week on one charge. In contrast, smartphones are almost always equipped with cameras, can use apps, and are able to access the internet and specifically social media. However, smartphone battery life is limited and AYAs reported needing to charge their smartphones daily or sometimes twice a day. These different functionalities were important to AYAs, and therefore, a substantial proportion of AYAs owned 2 phones. Across age and gender, making calls was the most consistently reported phone practice, followed by money transfer and social media. In terms of differences by gender and age, men and older AYAs (20-24 years) more often reported using their phones for work, and young people (15-19 years) more frequently reported playing games.

AYAs from a range of professions viewed phone ownership as a necessity for conducting business. This included boda boda drivers, a motorcycle taxi driver, and hairstylists. One woman who was about 30 years old and a hairstylist described as follows:

> *I use my phone to get customers, they call on it, or someone can send me a picture of hair style and she says that she wants the same style. I use the phone to know more different styles for brides because it is my job to style brides...*

For this hairstylist, the phone served as a critical tool for gathering information, such as *different styles for brides*, and for communicating with customers.

Barriers to Using Phones

As illustrated in quote 1 in Table 2, the most frequent barrier to using a phone was the lack of electricity for charging. As noted in quote 2 in Table 2, other common barriers were lack of network coverage, lack of airtime, and the overall costs of owning and maintaining a phone.

Table 2. Barriers to using phones quotes.

Quote number	Topic	Quote
Quote 1	Electricity	“You know in Uganda electricity is not constant and you are not certain that electricity will be available tomorrow. So, you have to charge every day, yet we pay a lot of money for electricity.”
Quote 2	Network and cost	“The battery is weak and it gets low quite often. Another thing is network, sometimes the network is not good. I have a friend in Kaliro but usually when you are talking it loses network and I do not get what he is talking. Then also money for airtime, buying sms, it is expensive for me because you have to load SMS for a day if you want to communicate with a friend.”
Quote 3	Unwanted messages	“Unwanted text messages. Haha. They are my biggest problem. I get stressed out and it ruins my day when I get them. Then there are frauds who lie to us on the phone that we have won something when they want to take our money. It makes me want to throw away my phone. Also, stalkers who refuse to introduce themselves and leave you in suspense. Then the problems with phones is that even if you break up with someone, they can't leave you alone.”

Three participants noted that strangers calling them were an added burden to using their phones, as described in quote 3 in Table 2. More women than men described unwanted messages and calls from the opposite sex as a major problem related to mobile phones. In addition, 2 participants reported that the OTT tax [26] was a barrier to using their phone.

Potential Perils of Mobile Phones

In response to the question *Do people ever get into trouble with their phones?*, the most common response was that sending explicit content could lead to trouble, as indicated in quote 4 in Table 3.

Table 3. Trouble with phones.

Quote number	Topic	Quote
Quote 4	Explicit content	"One man here in town shared his personal video clip with a sex worker on WhatsApp, and then she also shared it with others. This man was famous in town and he lost dignity because of this video clip."
Quote 5	Trouble with sexual partners	"Conflicts happen in marriage because people cheat a lot using phones. They are always in touch with other extra marital partners...I saw it on WhatsApp on my friend's phones. It was even within this town, some guy was touching himself on a video and sent it to the wife of his best friend and yet it's the woman's husband who sent a text to a phone number of some guy in his wife's phone. The husband pretending to be his wife asked him to send a video of himself nude and he did. When he sent the video, the guy leaked it. The guy got embarrassed."
Quote 6	Trouble with strangers	"A stranger may call you and ask you to meet them when they have other 'programs' and he may end up killing you. The other thing is when it comes to relationships, you can fall in love with someone you don't know and they cause you harm, like taking you for ritual sacrifice or raping or killing you."
Quote 7	Trouble with sexual partners and intimate partner violence	"I am a married woman. A phone can bring me problems if 'kojja' [uncle] that I work with admires me and asks me for my number. If I give it to him and he calls me when I am at home and he starts telling me that I look good among other things, my husband will become jealous, I will end up being beaten or we shall fight. He will start saying that I am promiscuous. The phone will have caused me problems."

The next most frequent *peril* noted was potential trouble related to sexual partners. In fact, all of the women and half of the men we interviewed reported that they faced trouble related to mobile phones and their sexual relationships. One woman's description of this trouble related to the themes of sexual partnership and sending explicit messages is included in quote 5 in Table 3.

Trouble related to mobile phones and sexual partners most often involved perceived infidelity, although in a few cases, this also included meeting new partners who were HIV positive and meeting new and physically dangerous partners, as described in quote 6 in Table 3. The ritual sacrifice noted in quote 6 may be more folktale rather than actual practice. However, intimate partner violence stemming from mobile phone communication was a concern for many women. For example, in quote 7 in Table 3, a 24-year-old woman described how calls to her mobile phone from other men stoked her husband's jealousy and led to intimate partner violence.

Communicating With Sexual Partners

Nearly every participant reported using their phone to communicate and flirt with their sexual partners. Across age and gender, the most common way to communicate with partners was through phone calls, followed by texting, and less frequently social media. In quote 8 in Table 4, a 21-year-old man succinctly replied to the question *How would you make plans to meet your partner?* Likewise, in quote 9 in Table 4, a 20-year-old man explained how he made plans with his partner and how these plans depended on her mobile phone. An 18-year-old woman described a similar pattern, "He calls me and tells me where to meet him."

A 20-year-old girl explained that despite the distance that separates her school from her partner's home, they use their mobile phones to make plans and meet up, as detailed in quote 10 in Table 4. Describing a similar process of using mobile phones to overcome geographical and logistical obstacles, an 18-year-old man spoke about communicating with his partner and arranging a meeting with a partner who lives in a different community in quote 11 in Table 4.

The following passage demonstrates how our EPO at a boda boda stage corroborates these findings. A boda boda stage is a place where motorcycle taxi drivers gather and wait for customers, and extensive research has documented the hierarchy and organizational structure of boda boda stages in Uganda [42]. This boda boda stage, constructed using timber and roofed with tarpaulin, was overcrowded with motorcycles parked close to each other as if they were displayed for sale and was located at a junction where dirt roads branched off the main paved road. Author 5 spoke with a talkative young man, who was the vice-chairperson or the second in command below the chairperson of the boda boda stage. Dressed in tight black jeans with a black T-shirt and holding 2 phones, a smartphone and a basic phone, the vice-chairperson explained, "Here is my smartphone, I bought it for UGX 350,000 (US $100), I use it for WhatsApp and Facebook." The vice-chairperson described how "most young people at boda boda stage use these smartphones purposely Okukwana Bakazi." When asked to clarify, the vice-chairperson explained okukwana means getting a woman and starting a relationship with her. He continued, "There is no way you can connect with a woman without a phone." Later, another boda boda driver supported this perspective and further explained that he used Facebook to "post messages using the smart phone. Sometimes these female sexual partners are in town." This driver also showed his basic phone that he said he used to communicate with customers.

Table 4. Mobile phones for communicating with sexual partners.

Quote number	Topic	Quote
Quote 8	Call, text, and WhatsApp	"I would say in three ways. I would call her. I would text her or would go to WhatsApp and send her an audio on WhatsApp."
Quote 9	Meeting my partner	"In most cases, I first call my sex partner to meet with her. I don't see any other plan I would consider meeting my sex partner without a phone call. If her mobile phone number is switched off, I must wait until it is switched on."
Quote 10	Back to school	"Now for example, we shall go back at school on 4th of February, for the first two weeks we are usually not busy because they are still registering and receiving students and we are waiting for results. So, over the weekend, I have the phone it is here (she shows it to the interviewer), I call him, or he calls me and tells me that I am going to come, and I also tell him that it is fine to come since we are mature."
Quote 11	A partner in a different community	"I call her wherever she is or I can send her a message and tell her that I want you to come and see me next week. She can give me her program and tell me that I will not be able to make it. If she accepts, I send her transport and she comes then we meet. In case she comes, she spends time with me on that day and then she goes back. I give her transport then she goes back. We keep communicating over the phone."

Mobile Phone–Specific Gift-Giving Practices

Nearly all the men in this study, but few women, reported giving gifts to their sexual partners. Given the breadth of the literature on transactional sex, we will focus our analysis on instances of mobile phone or mobile phone–related gift giving. Some men also reported giving their partner phones and airtime, but no women did so. After AYAs described gift-giving practices, we probed further and asked the gift givers if *they expected anything in return*. No participants indicated that they expected anything in return for these gifts.

Most participants did, however, report receiving gifts from their sexual partners. Many more women than men reported getting a phone or airtime as a gift from their sexual partner. It was also more common for older AYAs (20-24 years) than younger AYAs (15-19 years) to receive a phone. The practice of giving airtime was common for both older and younger AYAs. Author 5 noted how the vice-chairperson of the boda boda stage succinctly described one reason men buy phones for women, "It is true, the boda boda men buy phones for women to make it easier to communicate with them."

When we asked *Does your partner expect anything in return*?, most respondents replied *no*. However, a few women and 1 man who received gifts felt that their partner expected something in return. In terms of age groups, no older AYAs responded *yes* to this question; however, a few younger AYAs responded *yes*. One 18-year-old married woman responded that "For him, he did everything thinking that we must have sex." Another exchange with a young woman who was a student was as provided in Textbox 1.

Textbox 1. Exchange with a young woman who was a student.

Author 7: *Okay do you think that your partner expects you to give him something because he has given you a gift?*

AYA: *Mmmm...yes...*

Author 7: *Okay...so what do you think he expects in return?*

AYA: *Mmm...hahaha...the "other thing."*

Author 7: *Do you mean sex?*

AYA: *Yes that one hahaha.*

Author 7: *What makes you think he expects that?*

AYA: *I think it's because for all that he does for, that is the way I can pay him back.*

A young single man who was a student replied as follows:

> *Okay, I would expect a bigger gift; this is one way to test her. However, for the case of our young women, I know that they will always expect something after offering small gifts.*

Author 5, who conducted this interview, included in his field notes that the word *something* implied sexual relations. Both men and women considered gift giving to be a common practice with sexual partners. Interestingly, these responses suggest that some youths who received gifts believed that the gift giver expected sex in return for the gift. However, none of the AYAs who gave gifts reported that they expected sex in return for gifts they had given.

Promise of Mobile Phones

To gather information about how phones may present a key opportunity for gaining health information, we asked participants the open-ended question, *How do you get health information?* Participants most frequently noted getting health information from the Rakai Project and other health workers. Television and radio were also frequently mentioned as sources of health information. In the region, bars and other small shops often display a television, which becomes a public gathering space. One 20-year-old student described her sources of health information as consisting of radio and television, as described in quote 12 in Table 5.

Table 5. Mobile phones and health information.

Quote number	Topic	Quote
Quote 12	Radio and television	"In most cases there is a program on Capital FM Radio, it is there at 8:00 and there is a doctor I always listen to. There are also some health tips on the radio and TV. Every day I have to listen to that doctor."
Quote 13	Health worker call	"One health worker called me to come to the clinic and meet health workers. She told me to come and test for HIV. So, I left my work and went to the clinic."
Quote 14	Not for health	"No, we do not usually think of using phones to get health-related services. Using phones to get health related information? Another girl who is a customer echoes, no we use our phone to call our men (laughing)."
Quote 15	Need for information	"The health information that I would like to access on my phone is to be able to know the ways in which I can protect myself from HIV."
Quote 16	An emergency	"I fell sick, I did not have the energy to press the phone but during that time I was still staying with my wife. She brought the phone and I told her to check in my phone a health worker called, she called her and told her I am very sick, and I need quick medical attention. She asked her where I am at that time and my wife told her that I am at home, but I cannot walk. So, the health worker came from the health center to my home and gave me medication and the situation improved. However, if it was not for a phone, remember I could not walk and the wife was already in fear this could make me lose my life."

Only 1 participant mentioned that school was a place where they received health information, and 1 participant noted that using her phone to search the internet was a source of health information. When we probed and specifically asked if people had *ever used their mobile phones to gather health information*, a little less than half of the participants said they had. A few participants described how they had used their phones to stay in touch with a health worker, and a couple of other participants described searching the internet for health information, as detailed in quote 13 in Table 5. However, it appeared rare for people to use their phones for health purposes, as a couple of young women in a hair salon explained in quote 14 in Table 5.

Although currently phones are rarely used to access health information, most participants reported that they would like to use their phones for health-related information in the future. Most frequently, participants wanted to use their phones to access information about HIV prevention, as one young woman succinctly explained in quote 15 in Table 5. Following HIV, participants also frequently noted that mobile phones could be used for gathering information about other diseases, family planning, and finding a doctor.

Finally, most participants spoke of a time when they needed to use their phone in an emergency. Of these, the most common emergencies were health related, and a number of women also specifically described using their phones in an emergency related to childbirth and labor, whereas a few participants also described other types of medical emergencies. One young man explained how a mobile phone was critical during his incapacitation from malaria, as described in quote 16 in Table 5. Many participants felt that their mobile phones were a valuable resource during health-related emergencies.

Discussion

Principal Findings

In a region where many roads are unpaved and people have limited economic and health resources, mobile phones are changing the ways that young people form social connections [5,43-45], and these changing social dynamics have implications for sexual relationships [3]. This study demonstrates how mobile phone ownership is critical for engaging in economic, social, and sexual behaviors in the Rakai region. In fact, nearly every participant used a mobile phone to flirt and communicate with sexual partners. This finding provides some insight in terms of a possible mechanism for the quantitative associations between owning a mobile phone and having a greater number of sexual partners and between phone ownership and positive HIV status for young women [3]. Finally, although few AYAs currently use their phones for health-related purposes, we show that there may be great potential for mobile phones to serve as tools for health promotion in the future.

Potential Health-Related Perils

The psychological and social consequences of sending and receiving pornographic images and videos was one of the main perils of mobile phones cited by participants. In addition, for all of the women we interviewed, and many of the men, using a mobile phone caused trouble, such as issues of suspicion and jealousy, with their sexual partner. Women, in particular, reported that their male partners were suspicious of any phone use. For some women, the consequences of owning a phone outweighed the benefits, and these women chose to either get rid of their phones or conceal their phone ownership from their primary sexual partners.

It was also clear that mobile phones were frequently used to communicate with sexual partners. As one young man explained "there is no way you can connect with a woman without a phone." In addition, consistent with previous research, AYAs frequently noted the need to move to pursue jobs or education [46-48]. Owning a mobile phone enabled these AYAs to maintain social and sexual relationships despite the geographical distance imposed by job searches and attending boarding school.

In line with previous ethnographic research, we found mobile phones were at times a currency in transactional sex. This was a complex dynamic and should be interpreted in the context of previous work that has cautioned against assuming that transactional sex is specific to African contexts and noted that relationships where there is an exchange of money may also have important affective and social dimensions [11]. We found that when AYAs gave mobile phone–related gifts, they never

reported expecting anything in return. However, some AYAs who received gifts (eg, mobile phones or airtime) did *feel* there was an expectation that they engage in sexual behaviors after receiving a gift.

Promise of Mobile Phones for Health Promotion

Consistent with previous literature, we found that phones were critical for engaging in social and economic interactions [1,5,45]. From hairstylists to boda boda drivers, a mobile phone was required to conduct daily business, and this was more common for older participants as compared with younger participants. In contrast, more younger than older AYAs reported using their phones to play games. This may be relevant for future interventions that could leverage interest in mobile phone–based games to disseminate health information for younger AYAs.

Of note, few AYAs reported ever using their mobile phones for purposes that might promote their health. However, most participants expressed interest in using their phones for health-related purposes in the future. This disconnect highlights a critical area for future researchers to address the health needs of AYAs in southern Uganda and perhaps in other resource-limited settings. AYAs reported an interest in using their phones to gather information about diseases and family planning. In fact, AYAs in the region currently have limited access to sexual health information [4]. Half of all pregnancies in Uganda are unintended, and the vast majority of these pregnancies (88%) occur among adolescents [4].

AYAs overwhelmingly stated that they would like to use their mobile phones to stay up to date on information about HIV prevention. HIV prevention is certainly a crucial and pressing issue in the region. As noted earlier, East Africa has the greatest number of new HIV infections in the world, with AYAs at the greatest risk [9]. In fact, AIDS is the leading cause of death for AYAs in the region and the second leading cause of death for AYAs worldwide [49,50]. In Uganda, people under 30 years comprise 75% of the total population [21], and despite extensive combination prevention efforts, the HIV incidence rate for AYAs remains high [9]. AYAs in the region are highly mobile [46-48], and evidence suggests that structural factors force AYAs to move and search for work after they leave school [46]. Mobility makes it difficult to track AYAs for recommended health care [51], and mobility is also associated with a greater risk of HIV infection [47,48]. Accordingly, our analyses suggest that mobile phones present an opportunity to reach these highly mobile AYAs [5,43,44,46].

Despite the potential promise for future mobile phone–based health interventions, critical barriers remain. Most notably, AYAs in the region had inconsistent access to electricity. To ameliorate this obstacle, some AYAs maintain 2 phones—1 button phone that can last up to 1 week after a single charge and 1 smartphone that can access apps and the internet but requires daily charging. Access to electricity is also a potential barrier for researchers seeking to design interventions that would require smartphones or tablets.

Limitations

We believe that our KII coding achieved thematic saturation for the overall AYAs' sample. However, given our relatively small sample (N=31), we were only able to offer limited insights regarding health risk and health promotion opportunities by age group (eg, 15-19 years vs 20-24 years). Thus, a limitation of this study is that we may not have achieved saturation among these subgroup categories [37,40]. It is also possible that within our strata (owning a phone and not owning a phone), further data collection could yield additional insights with implications for mobile phone–related health risks and intervention opportunities in the region. For example, future research could focus exclusively on AYAs who own phones to examine specific mobile phone health intervention components and approaches that would best meet their needs and address barriers such as intermittent internet and electricity access.

Strengths

The use of purposive sampling from a population-based cohort allowed our research team to reliably and efficiently identify AYAs among the strata of focus (owning a phone and not owning a phone). Furthermore, by using multiple methods of data collection, KII and EPO, we were able to corroborate KII statements from the 3 strata through ecologically valid observations of behaviors in the community.

Conclusions

AYAs' developmental period is associated with the increased importance of peers, increased technology use, increased mobility, and initiation of sex [47,52,53]. A recent systematic review of digital innovations for HIV and sexually transmitted infections noted the need for research to tailor behavioral risk-reduction interventions to specific contexts and populations [13]. In southern Uganda, we found that technology use, mobility, and sexual behaviors were interrelated. Nearly all AYAs relied on mobile phones for making sexual partnerships, and in many instances, their phones allowed them to overcome significant geographic and logistical barriers to maintain long-distance sexual partnerships. Few AYAs used their phones to access health information, although most were interested in using their phones to access health information in the future, with specific enthusiasm for current information on HIV prevention. Given these factors, adolescence is a potentially critical developmental period for introducing preventive mobile phone–based interventions to provide AYAs in resource-limited settings with current and credible family planning and HIV prevention information and services.

Acknowledgments

The authors would like to thank the AYAs who shared their time and perspectives. This project was supported by a grant from the National Institute of Child Health and Human Development (R01HD091003, principal investigator: JS, MD, MPH). In addition, PK's contribution was also supported by grants from the National Institute of Mental Health (K01MH122319, principal investigator: PK, PhD; T32 MH019139, principal investigator: Theodorus Sandfort, PhD) and an award to the HIV Center for

Clinical and Behavioral Studies at the New York State Psychiatric Institute and Columbia University (P30MH43520, principal investigator: Robert Remien, PhD). AB's contribution was supported by the National Institute of Child Health and Human Development Predoctoral Fellowship in Gender, Sexuality, and Health (5T32HD049339, principal investigator: Constance A Nathanson, PhD), and the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1644869. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the National Institute of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key informant interview cell phone qualitative interview guide.
[DOCX File , 23 KB - jmir_v23i2e17837_app1.docx]

References

1. Poushter J, Oates R. Cell Phones in Africa: Communication Lifeline; Texting Most Common Activity, but Mobile Money Popular in Several Countries. Pew Research Center. 2015. URL: https://www.pewresearch.org/global/2015/04/15/cell-phones-in-africa-communication-lifeline/ [accessed 2020-12-23]
2. Silver L, Johnson C. Basic Mobile Phones More Common than Smartphones in Sub-Saharan Africa. Pew Research Center. 2019. URL: https://www.pewresearch.org/global/2018/10/09/majorities-in-sub-saharan-africa-own-mobile-phones-but-smartphone-adoption-is-modest/ [accessed 2020-12-22]
3. Nalugoda F, Kreniske P, Hofer S, Zhong X, Wei Y, Grilo SA, et al. Cell phones, sexual behaviors and HIV prevalence in Rakai, Uganda: a cross sectional analysis of longitudinal data. AIDS Behav 2020 May;24(5):1574-1584 [FREE Full text] [doi: 10.1007/s10461-019-02665-8] [Medline: 31520238]
4. In Uganda, Sexual and Reproductive Health Services and Information Fall Short for Adolescents. Guttmacher Institute. 2019 Mar 19. URL: https://www.guttmacher.org/news-release/2019/uganda-sexual-and-reproductive-health-services-and-information-fall-short [accessed 2020-12-22]
5. Porter G, Hampshire K, de Lannoy A, Bango A, Munthali A, Robson E, et al. Youth livelihoods in the cellphone era: perspectives from urban Africa. J Int Dev 2018 Jan 10;30(4):539-558 [FREE Full text] [doi: 10.1002/jid.3340]
6. Hampshire K, Porter G, Owusu SA, Mariwah S, Abane A, Robson E, et al. Informal m-health: How are young people using mobile phones to bridge healthcare gaps in Sub-Saharan Africa? Soc Sci Med 2015 Oct;142:90-99 [FREE Full text] [doi: 10.1016/j.socscimed.2015.07.033] [Medline: 26298645]
7. Archambault J. Cruising through uncertainty: Cell phones and the politics of display and disguise in Inhambane, Mozambique. American Ethnologist 2013;40(1):88-101 [FREE Full text] [doi: 10.1111/amet.12007]
8. Stark L. Transactional sex and mobile phones in a Tanzanian slum. Suomen Antropol 2013;38(1):12-36 [FREE Full text]
9. Grabowski MK, Serwadda DM, Gray RH, Nakigozi G, Kigozi G, Kagaayi J, et al. HIV prevention efforts and incidence of HIV in Uganda. N Engl J Med 2017 Nov 30;377(22):2154-2166 [FREE Full text] [doi: 10.1056/nejmoa1702150]
10. Kenny E. 'Phones mean lies': secrets, sexuality, and the subjectivity of mobile phones in Tanzania. Econ Anthropol 2016 Jun 7;3(2):254-265 [FREE Full text] [doi: 10.1002/sea2.12062]
11. Mojola SA. Love, Money, and HIV: Becoming a Modern African Woman in the Age of AIDS. California, USA: University of California Press; 2014:1-6.
12. Ippoliti NB, L'Engle K. Meet us on the phone: mobile phone programs for adolescent sexual and reproductive health in low-to-middle income countries. Reprod Health 2017 Jan 17;14(1):11 [FREE Full text] [doi: 10.1186/s12978-016-0276-z] [Medline: 28095855]
13. Daher J, Vijh R, Linthwaite B, Dave S, Kim J, Dheda K, et al. Do digital innovations for HIV and sexually transmitted infections work? Results from a systematic review (1996-2017). BMJ Open 2017 Nov 3;7(11):e017604 [FREE Full text] [doi: 10.1136/bmjopen-2017-017604]
14. Shah R, Watson J, Free C. A systematic review and meta-analysis in the effectiveness of mobile phone interventions used to improve adherence to antiretroviral therapy in HIV infection. BMC Public Health 2019 Jul 9;19(1):- [FREE Full text] [doi: 10.1186/s12889-019-6899-6]
15. Wang Z, Zhu Y, Cui L, Qu B. Electronic Health Interventions to Improve Adherence to Antiretroviral Therapy in People Living With HIV: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2019 Oct 16;7(10):e14404 [FREE Full text] [doi: 10.2196/14404] [Medline: 31621641]
16. Musiimenta A, Tumuhimbise W, Atukunda E, Mugaba AT, Muzoora C, Armstrong-Hough M, et al. Mobile health technologies may be acceptable tools for providing social support to tuberculosis patients in rural Uganda: a parallel mixed-method study. Tuberc Res Treat 2020 Jan 7;2020:- [FREE Full text] [doi: 10.1155/2020/7401045] [Medline: 31969998]

17. Musiimenta A, Tumuhimbise W, Mugaba AT, Muzoora C, Armstrong-Hough M, Bangsberg D, et al. Digital monitoring technologies could enhance tuberculosis medication adherence in Uganda: Mixed methods study. Journal of Clinical Tuberculosis and Other Mycobacterial Diseases 2019 Dec;17:100119 [FREE Full text] [doi: 10.1016/j.jctube.2019.100119]
18. Ggita JM, Ojok C, Meyer AJ, Farr K, Shete PB, Ochom E, et al. Patterns of usage and preferences of users for tuberculosis-related text messages and voice calls in Uganda. Int I Iuberc Lung Dis 2018 May 01;22(5):530-536 [FREE Full text] [doi: 10.5588/ijtld.17.0521]
19. Ggita JM, Katahoire A, Meyer AJ, Nansubuga E, Nalugwa T, Turimumahoro P, et al. Experiences and intentions of Ugandan household tuberculosis contacts receiving test results via text message: an exploratory study. BMC Public Health 2020 Mar 12;20(1):- [FREE Full text] [doi: 10.1186/s12889-020-8427-0]
20. Chang LW, Grabowski MK, Ssekubugu R, Nalugoda F, Kigozi G, Nantume B, et al. Heterogeneity of the HIV epidemic in agrarian, trading, and fishing communities in Rakai, Uganda: an observational epidemiological study. The Lancet HIV 2016 Aug;3(8):e388-e396 [FREE Full text] [doi: 10.1016/s2352-3018(16)30034-0]
21. Uganda Bureau of Statistics. National Population and Housing Census 2014. Uganda Bureau of Statistics. 2016. URL: https://unstats.un.org/unsd/demographic/sources/census/wphc/Uganda/UGA-2016-05-23.pdf [accessed 2020-12-23]
22. Uganda Bureau of Statistics. Uganda National Household Survey 2016/2017 Report. Uganda Bureau of Statistics. Kampala: Uganda Bureau of Statistics; 2020. URL: https://www.ubos.org/wp-content/uploads/publications/03_20182016_UNHS_FINAL_REPORT.pdf [accessed 2020-12-23]
23. Gillwald A, Mothobi O, Ndiwalana A, Tusubira T. The State of ICT in Uganda. Research ICT Africa. 2019. URL: https://researchictafrica.net/publication/the-state-of-ict-in-uganda/ [accessed 2020-12-23]
24. National Labour Force Survey 2016/17. Uganda Bureau of Statistics. 2018. URL: https://www.ubos.org/wp-content/uploads/publications/10_2018Report_national_labour_force_survey_2016_17.pdf [accessed 2020-12-22]
25. National Information Technology Survey 2017/18 Report. NITA Uganda. Kampala: NITA Uganda; 2018 Mar. URL: https://www.nita.go.ug/sites/default/files/publications/National%20IT%20Survey%20April%2010th.pdf [accessed 2020-12-22]
26. Ratcliffe R, Okiror S. Millions of Ugandans quit internet services as social media tax takes effect. The Guardian. URL: https://www.theguardian.com/global-development/2019/feb/27/millions-of-ugandans-quit-internet-after-introduction-of-social-media-tax-free-speech [accessed 2020-12-22]
27. James J. Patterns of Mobile Phone Use in Developing Countries: Evidence from Africa. Soc Indic Res 2013 Nov 30;119(2):687-704 [FREE Full text] [doi: 10.1007/s11205-013-0510-9]
28. Orotin P, Quisenbery W, Ted S. A Study on Factors Facilitating Access to Mobile Phone Money in Uganda. GJBMS 2013 Aug 20;3(6):279-291 [FREE Full text] [doi: 10.15580/gjbms.2013.6.071213722]
29. Wawer MJ, Sewankambo NK, Serwadda D, Quinn TC, Kiwanuka N, Li C, et al. Control of sexually transmitted diseases for AIDS prevention in Uganda: a randomised community trial. The Lancet 1999 Feb;353(9152):525-535 [FREE Full text] [doi: 10.1016/s0140-6736(98)06439-3]
30. Daiute C. Narrative Inquiry: A Dynamic Approach. In: Qualitative Methods. California: SAGE Publications, Inc; Oct 2013:1-320.
31. Wertsch JV. Narrative Tools of History and Identity. Culture & Psychology 2016 Jul 24;3(1):5-20 [FREE Full text] [doi: 10.1177/1354067x9700300101]
32. Bruner J. Actual Minds, Possible Worlds. In: THE JERUSALEM-HARVARD LECTURES. Cambridge: Harvard University Press; Oct 15, 1987:1-222.
33. Charmaz K. Constructing Grounded Theory. In: Introducing Qualitative Methods series. : Thousand Oaks, California: Sage Publication; Mar 2014:1-416.
34. Garcia J, Parker C, Parker RG, Wilson PA, Philbin MM, Hirsch JS. 'You're Really Gonna Kick Us All Out?' Sustaining Safe Spaces for Community-Based HIV Prevention and Control among Black Men Who Have Sex with Men. PLoS ONE 2015 Oct 22;10(10):e0141326 [FREE Full text] [doi: 10.1371/journal.pone.0141326]
35. Hirsch J, Wardlow H, Phinney H. ‘No one saw us’: reputation as an axis of sexual identity 1. In: Understanding Global Sexualities. London: Informa UK Limited; 2012:1-17.
36. Emerson R, Fretz R, Shaw L. Writing Ethnographic Fieldnotes. In: The University of Chicago Press Books. Chicago: University of Chicago; 2011.
37. Urquhart C. Grounded Theory for Qualitative Research: A Practical Guide. In: Sage Research Methods. California: SAGE Journals; 2013:A.
38. Creswell J, Poth C. Customers who viewed Qualitative Inquiry and Research Design: Choosing Among Five Approaches. California: SAGE Journals; 2018.
39. Web Application for Managing Analyzing, Presenting QualitativeMixed Method Research Data. Web Application for Managing Analyzing, and Presenting Qualitative and Mixed Method Research Data. - 2018:- [FREE Full text] [doi: 10.2196/preprints.13265]
40. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018 Sep 14;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]

41. Shapka JD, Onditi HZ, Collie RJ, Lapidot-Lefler N. Cyberbullying and Cybervictimization Within a Cross-Cultural Context: A Study of Canadian and Tanzanian Adolescents. Child Dev 2018 Jan 19;89(1):89-99 [FREE Full text] [doi: 10.1111/cdev.12829] [Medline: 28523643]
42. Doherty J. Life (and limb) in the fast-lane: disposable people as infrastructure in Kampala's boda boda industry. Critical African Studies 2017 May 03;9(2):192-209 [FREE Full text] [doi: 10.1080/21681392.2017.1317457]
43. Porter G, Hampshire K, Abane A, Munthali A, Robson E, Bango A, et al. Intergenerational relations and the power of the cell phone: Perspectives on young people's phone usage in sub-Saharan Africa. Geoforum 2015 Aug;64:37-46 [FREE Full text] [doi: 10.1016/j.geoforum.2015.06.002]
44. Porter G. Mobilities in Rural Africa: New Connections, New Challenges. Annals of the American Association of Geographers 2016 Jan 04:1-8 [FREE Full text] [doi: 10.1080/00045608.2015.1100056]
45. Porter G, Hampshire K, Abane A, Munthali A, Robson E, Tanle A, et al. Connecting with home, keeping in touch: physical and virtual mobility across stretched families in sub-Saharan Africa. Cambridge Core 2018 May 24;88(2):404-424 [FREE Full text] [doi: 10.1017/s0001972017000973]
46. Kreniske P, Grilo S, Nakyanjo N, Nalugoda F, Wolfe J, Santelli JS. Narrating the Transition to Adulthood for Youth in Uganda: Leaving School, Mobility, Risky Occupations, and HIV. Health Educ Behav 2019 Feb 21;46(4):550-558 [FREE Full text] [doi: 10.1177/1090198119829197]
47. Olawore O, Tobian AAR, Kagaayi J, Bazaale JM, Nantume B, Kigozi G, et al. Migration and risk of HIV acquisition in Rakai, Uganda: a population-based cohort study. The Lancet HIV 2018 Apr;5(4):e181-e189 [FREE Full text] [doi: 10.1016/s2352-3018(18)30009-2]
48. Schuyler AC, Edelstein ZR, Mathur S, Sekasanvu J, Nalugoda F, Gray R, et al. Mobility among youth in Rakai, Uganda: Trends, characteristics, and associations with behavioural risk factors for HIV. Global Public Health 2015 Aug 27;12(8):1033-1050 [FREE Full text] [doi: 10.1080/17441692.2015.1074715]
49. Miles to Go: Closing Gaps, Breaking Barriers, Righting Injustices. UNAIDS. URL: http://www.unaids.org/sites/default/files/media_asset/miles-to-go_en.pdf [accessed 2020-12-22]
50. HIV and Youth: Maternal, newborn, child and adolescent health. World Health Organization. URL: https://www.who.int/maternal_child_adolescent/topics/adolescence/hiv/en/ [accessed 2020-12-22]
51. Billioux VG, Chang LW, Reynolds SJ, Nakigozi G, Ssekasanvu J, Grabowski MK, et al. Human immunodeficiency virus care cascade among sub-populations in Rakai, Uganda: an observational study. Journal of the International AIDS Society 2017 Jun 05;20(1):21590. [doi: 10.7448/ias.20.1.21590]
52. Breakwell GM, Breakwell GM. Risk Estimation and Sexual Behaviour. J Health Psychol 2016 Jul;1(1):79-91 [FREE Full text] [doi: 10.1177/135910539600100107]
53. Anderson KG, Beutel AM, Maughan-Brown B. HIV Risk Perceptions and First Sexual Intercourse Among Youth in Cape Town South Africa. IFPP 2007 Sep;33(03):0598-0105 [FREE Full text] [doi: 10.1363/3309807]

Abbreviations

ART: antiretroviral therapy
AYA: adolescents and young adult
EPO: ethnographic participant observation
KII: key informant interview
LMIC: low-to-middle-income countries
RCCS: Rakai Community Cohort Study
SBS: social and behavioral sciences

Edited by G Eysenbach; submitted 16.01.20; peer-reviewed by A Meyer, J Haberer; comments to author 14.02.20; revised version received 18.05.20; accepted 23.11.20; published 02.02.21.

Please cite as:
Kreniske P, Basmajian A, Nakyanjo N, Ddaaki W, Isabirye D, Ssekyewa C, Nakubulwa R, Hirsch JS, Deisher A, Nalugoda F, Chang LW, Santelli JS
The Promise and Peril of Mobile Phones for Youth in Rural Uganda: Multimethod Study of Implications for Health and HIV
J Med Internet Res 2021;23(2):e17837
URL: https://www.jmir.org/2021/2/e17837
doi:10.2196/17837
PMID:33528375

Original Paper

Sociodemographic, Health and Lifestyle, Sampling, and Mental Health Determinants of 24-Hour Motor Activity Patterns: Observational Study

Sonia Difrancesco[1], MSc; Harriëtte Riese[2], PhD; Kathleen R Merikangas[3], PhD; Haochang Shou[4], PhD; Vadim Zipunnikov[5], PhD; Niki Antypa[6], PhD; Albert M van Hemert[7], MD, PhD; Robert A Schoevers[2], MD, PhD; Brenda W J H Penninx[1], PhD; Femke Lamers[1], PhD

[1]Amsterdam Public Health Research Institute, Department of Psychiatry, Amsterdam UMC, Vrije Universiteit, Amsterdam, Netherlands

[2]Interdisciplinary Center Psychopathology and Emotion Regulation, Department of Psychiatry, Universitair Medisch Centrum Groningen, University of Groningen, Groningen, Netherlands

[3]Genetic Epidemiology Branch, Intramural Research Program, National Institute of Mental Health, Bethesda, MD, United States

[4]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, United States

[5]Department of Biostatistics, Johns Hopkins University, Baltimore, MD, United States

[6]Department of Clinical Psychology, Institute of Psychology, Leiden University, Leiden, Netherlands

[7]Department of Psychiatry, Leiden University Medical Center, Leiden, Netherlands

Corresponding Author:
Sonia Difrancesco, MSc
Amsterdam Public Health Research Institute
Department of Psychiatry
Amsterdam UMC, Vrije Universiteit
Oldenaller 1
Amsterdam, 1078XL
Netherlands
Phone: 31 643193730
Email: s.difrancesco@ggzingeest.nl

Abstract

Background: Analyzing actigraphy data using standard circadian parametric models and aggregated nonparametric indices may obscure temporal information that may be a hallmark of the circadian impairment in psychiatric disorders. Functional data analysis (FDA) may overcome such limitations by fully exploiting the richness of actigraphy data and revealing important relationships with mental health outcomes. To our knowledge, no studies have extensively used FDA to study the relationship between sociodemographic, health and lifestyle, sampling, and psychiatric clinical characteristics and daily motor activity patterns assessed with actigraphy in a sample of individuals with and without depression/anxiety.

Objective: We aimed to study the association between daily motor activity patterns assessed via actigraphy and (1) sociodemographic, health and lifestyle, and sampling factors, and (2) psychiatric clinical characteristics (ie, presence and severity of depression/anxiety disorders).

Methods: We obtained 14-day continuous actigraphy data from 359 participants from the Netherlands Study of Depression and Anxiety with current (n=93), remitted (n=176), or no (n=90) depression/anxiety diagnosis, based on the criteria of the Diagnostic and Statistical Manual of Mental Disorders, fourth edition. Associations between patterns of daily motor activity, quantified via functional principal component analysis (fPCA), and sociodemographic, health and lifestyle, sampling, and psychiatric clinical characteristics were assessed using generalized estimating equation regressions. For exploratory purposes, function-on-scalar regression (FoSR) was applied to quantify the time-varying association of sociodemographic, health and lifestyle, sampling, and psychiatric clinical characteristics on daily motor activity.

Results: Four components of daily activity patterns captured 77.4% of the variability in the data: overall daily activity level (fPCA1, 34.3% variability), early versus late morning activity (fPCA2, 16.5% variability), biphasic versus monophasic activity (fPCA3, 14.8% variability), and early versus late biphasic activity (fPCA4, 11.8% variability). A low overall daily activity level was associated with a number of sociodemographic, health and lifestyle, and psychopathology variables: older age (P<.001),

higher education level (P=.005), higher BMI (P=.009), greater number of chronic diseases (P=.02), greater number of cigarettes smoked per day (P=.02), current depressive and/or anxiety disorders (P=.05), and greater severity of depressive symptoms (P<.001). A high overall daily activity level was associated with work/school days (P=.02) and summer (reference: winter; P=.03). Earlier morning activity was associated with older age (P=.02), having a partner (P=.009), work/school days (P<.001), and autumn and spring (reference: winter; P=.02 and P<.001, respectively). Monophasic activity was associated with older age (P=.005). Biphasic activity was associated with work/school days (P<.001) and summer (reference: winter; P<.001). Earlier biphasic activity was associated with older age (P=.005), work/school days (P<.001), and spring and summer (reference: winter; P<.001 and P=.005, respectively). In FoSR analyses, age, work/school days, and season were the main determinants having a time-varying association with daily motor activity (all P<.05).

Conclusions: Features of daily motor activity extracted with fPCA reflect commonly studied factors such as the intensity of daily activity and preference for morningness/eveningness. The presence and severity of depression/anxiety disorders were found to be associated mainly with a lower overall activity pattern but not with the time of the activity. Age, work/school days, and season were the variables most strongly associated with patterns and time of activity, and thus future epidemiological studies on motor activity in depression/anxiety should take these variables into account.

(J Med Internet Res 2021;23(2):e20700) doi:10.2196/20700

KEYWORDS

actigraphy; functional data analysis; mental health; well-being; activity

Introduction

The near ubiquitous use of accelerometers in electronic devices ranging from smartphones to wrist-worn wearables provides the biomedical community with a potential richness of data that is useful for the study of health outcomes. Wrist-worn accelerometers have been used for more than 20 years by sleep researchers to estimate sleep and circadian activity rhythms [1,2], as well as by those studying patterns of physical activity [3]. Research indicates that disruptions in circadian activity rhythms, especially daily motor activity patterns, correlate with poor mental [4] and physical [5] health. Burton et al [6] and our recent results showed that a low level of daily motor activity is associated with depressive [7,8] and anxiety [7] disorders. In addition, sociodemographic and lifestyle factors, especially age and BMI, have been linked to disruptions in daily motor activity patterns. Compared with younger persons, older persons have been found to have lower motor activity patterns [9] and earlier bed and rise times, also known as early chronotype [10]. A higher BMI has been associated with a lower daytime activity level and a higher night-time activity level [11,12]. Circadian activity rhythms are also controlled externally by environmental and social cues. For instance, light is an important synchronizer for circadian activity rhythms [13] and has been shown to be effective in the treatment of sleep disorders [14] and affective disorders [15].

Despite the great interest in daily motor activity patterns and their association with health outcomes and other health factors, commonly used methodologies to analyze actigraphy data are limited in the description of circadian rhythms. Often used validated methods aggregate data in daily indices [16], which loses important information that may be a hallmark of circadian impairment. The traditional approach to actigraphy data analysis employs cosinor (ie, based on the mathematical formula of a cosine wave) or modified cosinor analyses that yield information concerning the amplitude of activity, the timing of "peak" activity, and the goodness-of-fit (ie, how close the pattern is to a cosine wave) [16]. Although they are often quite useful in people with robust activity patterns, these analyses assume the presence of a particular shape of activity (ie, a predictable pattern, such as a cosine waveform) that may be different in individuals with physical or psychological impairments. Functional data analysis (FDA) can be used to model the complete time series of actigraphy data with less restrictive assumptions [17]. Recent studies using functional principal component analysis (fPCA), an FDA technique, have shown that patterns of daily motor activity with specific shapes are associated with psychiatric clinical characteristics (ie, apathy [18], depressive and anxiety symptoms [19], and objectively assessed sleep [19]) in individuals with Alzheimer disease. Another FDA technique that is increasingly being used is function-on-scalar regression (FoSR), which analyzes the relative time-varying associations between each variable of interest and the activity patterns. In addition, this method yields valuable information about the time intervals in which the variables of interest have the greatest influence on activity patterns [20,21]. Banihashemi et al [9] have suggested that older age and higher BMI are linked to lower daytime activity levels and higher BMI and greater symptom severity are associated with nocturnal activity patterns suggestive of sleep disturbances in a population with affective disorders. Those findings are based on first attempts using FDA, and such approaches have not yet been explored in a population with depressive and anxiety disorders. In addition, no studies have extensively assessed the association between actigraphy functional curves and sociodemographic, lifestyle, and sampling factors (eg, season and work/school days). FDA may better capture the complexity and dynamics of daily motor activity to reveal important behavioral biomarkers. This may help us to understand whether intervening on circadian rhythms or sleep (eg, by light therapy or sleep intervention) could be a useful regimen to reduce depressive and anxiety disorders.

In this study, we examined the association between daily activity patterns, assessed using actigraphy and FDA, and (1) sociodemographic (ie, age, sex, partner status, and education level), health and lifestyle (ie, drinking, smoking, chronic

diseases, and BMI), and sampling (ie, season, and work/school day versus nonwork/nonschool day) factors, and (2) psychiatric clinical characteristics (ie, current/remitted depressive and anxiety disorders, severity of depressive, and medication use).

Methods

Sample

Participants from the Netherlands Study of Depression and Anxiety (NESDA) were selected to enroll in the Ecological Momentary Assessment (EMA) and Actigraphy substudy (NESDA-EMAA). NESDA is one of the cores sites of the Motor Activity Research Consortium for Health (mMARCH) [22,23], a collaborative network for the application of objective assessment of motor activity, sleep, and mood in population and clinical samples. Details about NESDA have previously been discussed extensively [24]. NESDA was designed to investigate the course of depressive and anxiety disorders over a period of several years and the factors that influence the development and prognosis of such disorders. NESDA participants were initially recruited for baseline assessment between 2004 and 2007 (n=2981) and seen for the fifth time at the 9-year follow-up assessment (2014-2017; n=2069) for a regular follow-up interview, including a psychiatric diagnostic interview. In total, 1776 individuals participated in face-to-face interviews. A total of 367 siblings of NESDA participants who met diagnostic criteria for a depressive or anxiety disorder and had the same biological parents as their sibling(s) were included as participants to NESDA's 9-year follow-up assessment. At the 9-year follow-up, we conducted the EMAA substudy among 384 participants. The NESDA study, including the EMAA component, was approved by the VUmc ethical committee (reference number 2003/183), and all participants gave informed consent for both the regular interview and the EMAA component. A flowchart of the NESDA-EMAA study was previously provided in Difrancesco et al [7]. Eligibility criteria included the following: (1) had a smartphone or were willing to use a smartphone provided by the study, (2) were willing to wear a wrist-worn actigraphy device, and (3) could be enrolled within 1 month of the NESDA interview. Siblings were invited if they did not have a current or past diagnosis of a depressive and/or anxiety disorder or another severe psychiatric disorder (such as psychotic or severe addiction disorder). Participants of the EMAA substudy were provided with a wrist-worn GENEActiv device (Activinsights Ltd) and wore it for 2 weeks on their nondominant wrist. The devices were initialized and set to collect raw activity measures at a frequency of 30 Hz. They also completed questions on their current mood states using EMA [25]. In this paper, we only report on the actigraphy component of the study. Of the 384 participants included in the NESDA-EMAA study, 14 had no available actigraphy data for several reasons, such as technical failure (see [7] for more details), resulting in 370 (96.4%) participants with available data. According to previously published criteria [26], participants' actigraphy data were included in analyses if at least 1 weekday and 1 weekend day of usable data were available, with at least 16 hours recorded per day and per night. The final sample consisted of 359 (359/384, 93.5%) participants with 13.68 (SD 1.26) valid days; 90.0% (323/359) of participants completed the protocol for 14 days.

Assessment of Sociodemographic, Health and Lifestyle, and Sampling Factors

Sociodemographic and health and lifestyle factors were assessed at the 9-year follow-up. Sociodemographic factors included age, gender, education level (expressed in years), and partner status. Health and lifestyle factors included BMI (kg/m^2), number of self-reported chronic diseases under treatment (eg, heart disease, diabetes, stroke, lung disease, osteoarthritis, cancer, ulcer, intestinal problems, liver disease, epilepsy, and thyroid gland disease), number of cigarettes smoked per day, and number of alcoholic drinks consumed per day. Sampling factors were assessed at the 9-year follow-up based on EMA and actigraphy assessment. Sampling factors included whether the actigraphy assessment was performed on a work/school day and the season in which the assessment was performed. Work/school day was identified with information from the EMA assessment. Season was determined based on the date of the actigraphy assessment (eg, 25/09/yyyy=autumn), and winter was used as reference.

Assessment of Depressive and/or Anxiety Disorders and Clinical Characteristics

As in the previous assessment periods, specially trained clinical research staff conducted the diagnostic interviews at the 9-year follow-up. The Composite International Diagnostic Interview (CIDI; version 2.1) [27] was used to establish diagnoses of depressive disorders (dysthymia and major depressive disorder) and anxiety (social anxiety disorder, panic disorder with and without agoraphobia, agoraphobia, and generalized anxiety disorder) based on the Diagnostic and Statistical Manual of Mental Disorders, fourth edition. For this study, we divided participants into three groups: (1) no lifetime depressive and/or anxiety disorders, (2) remitted depressive and/or anxiety disorders, defined as having a lifetime but not current (6-month) diagnosis, and (3) current depressive or anxiety disorder diagnosed in the past 6 months.

Clinical characteristics that were examined were severity of depressive symptoms and medication use (ie, antidepressant and benzodiazepine use). Severity of depressive symptoms was assessed with the 30-item Inventory of Depressive Symptomatology (IDS) [28]. Antidepressant and benzodiazepine use was based on drug container inspection, and medications were coded according to the Anatomical Therapeutic Chemical (ATC) classification of the World Health Organization. Antidepressant and/or benzodiazepine use was considered present if participants reported using them more than 50% of the time. Antidepressants included selective serotonin reuptake inhibitors (ATC code N06AB), tricyclic antidepressants (ATC code N06AA), and other antidepressants (ATC codes N06AF, N06AG, and N06AX); benzodiazepines included ATC codes N03AE, N05BA, N05CD, and N05CF.

Statistical Analyses

Descriptive Analyses

Distributions of all variables were checked on normality with Q-Q plots. For descriptive statistics, participants' sociodemographic, health and lifestyle, and sampling factors and clinical characteristics were compared between the three groups (ie, no, remitted, and current depressive and/or anxiety disorders). For normally distributed continuous data, analysis of variance tests were used, and for data with nonnormal distributions, Kruskal-Wallis tests were used. Chi-square tests were used to test differences in frequencies across the three groups. All analyses were performed with the statistical software R (version 1.0.143), and a *P* value of <.05 was considered statistically significant.

Assessment of Circadian Rhythm Patterns With fPCA

Raw actigraphy data were processed with open source R package GGIR (version 1.5-18) [29] according to published methods [30,31]. Processing of data included autocalibration, nonwear detection, identification of potentially corrupted data, collapsing of raw data to epoch level, and computation of missing data. Collapsing of raw data to epoch level was done by averaging 5-second data. Minute-to-minute daily actigraphy data were derived per participant by summing these 5-second data; day was defined as the 24-hour clock time interval. As days with at least 16 valid hours were included in the analyses, missing data points were replaced with the participant's data from the same time of day, averaged across the other valid days, to provide a person-specific informed approach.

Daily motor activity patterns were derived from minute-level actigraphy data with the R package fda for FDA (version 2.4.8) [32]. First, participants' minute-to-minute actigraphy data was pre-smoothed as linear combinations of a set of nine Fourier basis functions to capture the major trends in daily motor activity (ie, the same procedure as was applied by Zeitzer et al [19] and Gershon et al [33]). Second, fPCA was used to capture the principal directions of daily variation and dimension reduction. fPCA summarized the daily-specific features as the coordinates (called principal component scores) of participants' curves in the basis spanned by the principal components. The first four daily-specific features, referred to as functional principal components in this paper, were considered because together they explained at least 75% of data variability (a similar cutoff was previously used [33]).

The association between the extracted daily functional principal components for each participant and the participant's sociodemographic, health and lifestyle, and sampling factors and clinical characteristics were tested by using multiple generalized estimating equation (GEE) regressions, with each functional principal component as the outcome. GEE regression was used to account for correlations between repeated days per person. Separate models were run for each clinical characteristic (ie, current/remitted depressive and/or anxiety disorders, IDS score, and antidepressant use) and each model was adjusted for sociodemographic, health and lifestyle, and sampling factors; this was done to avoid collinearity induced by the high correlation between psychiatric clinical characteristics. Multiple testing correction was applied by controlling the false discovery rate [34].

Assessment of Time-Varying Associations of the Activity With Sociodemographic, Health and Lifestyle, and Mental Health Determinants With FoSR

Minute-to-minute daily actigraphy data were aggregated over 10 minutes and averaged over the assessment period for each participant (ie, similar procedure as was used by Goldsmith et al [21]). FoSR was used to study the time-varying associations between participants' sociodemographic, health and lifestyle, and sampling factors and clinical characteristics, with actigraphy data as the outcome (ie, to study the time-varying associations between several factors and activity patterns). For exploratory purposes, FoSR analysis was repeated for each clinical characteristic to account for the high correlation between psychiatric clinical variables. Each model was adjusted for sociodemographic, health and lifestyle, and sampling factors. Data were analyzed with the R script developed by Goldsmith [35].

Results

The sociodemographic, health and lifestyle, and sampling factors and clinical characteristics of the study participants are shown in Table 1. Of the total sample (N=359), 93 and 176 participants, respectively, had current and remitted depressive and/or anxiety disorders, while 90 participants had no current depressive and/or anxiety disorders. The current depressive/anxiety disorder group was heterogeneous in that 38.7% (36/93) had anxiety disorders only, 33.3% (31/93) had depressive and anxiety disorders, and 28.0% (26/93) had depressive disorders only. As expected, individuals with current depressive and/or anxiety disorders scored significantly higher on depressive symptoms (*P*<.001) and used antidepressants more frequently than individuals in the other groups, although no significant differences were found for benzodiazepine use.

Table 1. Participants' sociodemographics, health and lifestyle factors, actigraphy assessment and psychiatric characteristics, and medication use (N=359).

Variables	Current depressive and/or anxiety disorders (n=93)	Remitted depressive and/or anxiety disorders (n=176)	No depressive and/or anxiety disorders (n=90)	*P* value
Sociodemographics				
Age, mean (SD)[a]	50.1 (11.1)	48.2 (13.4)	51.3 (12.5)	.13
Female, n (%)[b]	58 (62.4)	120 (68.2)	50 (55.6)	.12
Education level (years), mean (SD)[a]	12.5 (3.4)	12.7 (2.8)	13.9 (2.9)	<.001
Has a partner, n (%)[b]	45 (48.4)	90 (51.1)	51 (56.7)	.52
Lifestyle factors [a]				
BMI, mean (SD)	27.1 (5.1)	26.6 (5.2)	26 (5.4)	.37
Number of chronic diseases, mean (SD)	1.1 (1.2)	1.0 (1.1)	0.6 (0.8)	.008
Number of cigarettes per day, mean (SD)	3.2 (6.5)	2.9 (6.1)	0.8 (2.7)	.004
Number of drinks per day, mean (SD)	0.5 (0.8)	0.7 (1.2)	0.8 (0.8)	.16
Actigraphy assessment characteristics				
Number of actigraphy days, mean (SD)[a]	13.7 (1.0)	13.6 (1.5)	13.8 (0.7)	.48
Measures on work/school days, n (%)[b]	398 (34.1)	769 (37.2)	467 (42.8)	<.001
Season of actigraphy measurement [b]				<.001
Winter, n (%)	277 (21.8)	701 (29.3)	369 (29.7)	
Autumn, n (%)	376 (29.5)	584 (24.4)	362 (29.1)	
Spring, n (%)	432 (33.9)	619 (25.8)	260 (20.9)	
Summer, n (%)	188 (14.8)	492 (20.5)	252 (20.3)	
Psychopathology				
Only depressive disorders, n (%)	26 (28.0)	46 (26.1)	N/A[c]	
Only anxiety disorders, n (%)	36 (38.7)	24 (13.6)	N/A	
Depressive and anxiety disorders, n (%)	31 (33.3)	106 (60.2)	N/A	
Inventory of Depressive Symptomatology, mean score (SD)[a]	28.6 (11.4)	20.9 (12.5)	6.0 (4.9)	<.001
Medication use				
Antidepressant users, n (%)[b]	35 (37.2)	34 (19.3)	2 (2.2)	<.001
Benzodiazepine users, n (%)[b]	5 (5.3)	8 (4.5)	0 (0.0)	.10

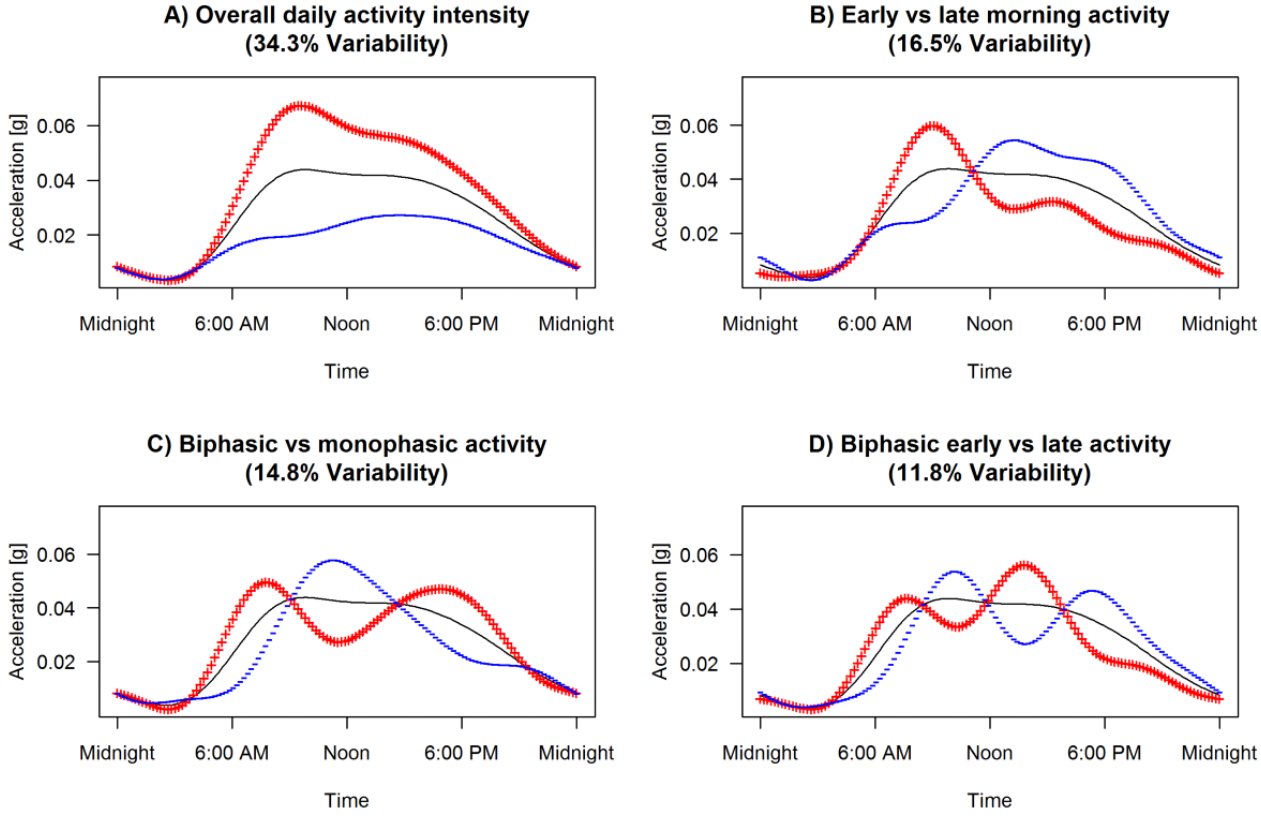
[a]Analysis of variance tests were performed.

[b]Chi-square tests were performed.

[c]N/A: not applicable.

Four components of daily activity patterns that explained 77.4% of the variability in the data were extracted using fPCA (Figure 1) and interpreted as follows. The first component was the overall daily activity level. The second component described early versus late morning activity and could be indicative of chronotype. The third component showed a biphasic versus monophasic activity pattern, while the fourth component represented an early versus late biphasic activity pattern. The biphasic pattern showed two cycles of increased activity with subsequent decreased activity.

Figure 1. Patterns of daily activity explaining 77.4% of variability in the data (N=359). The black line represents the average daily activity, the red line represents the average daily activity plus 1 SD of the functional principal component analysis (fPCA) score, and the blue line represents the average daily activity minus 1 SD of the fPCA score. (A) High (+) versus low (–) overall daily activity intensity. (B) Early (+) versus late (–) morning activity. (C) Biphasic (+) versus monophasic (–) activity. (D) Early (+) versus late (–) biphasic activity. Determinants associated with a pattern marked with (+) have a positive β value; if not, they have a negative β value.



A low overall daily activity level was associated with a number of sociodemographic, health and lifestyle, and psychopathology variables (Table 2): older age (P<.001), higher education level (P=.005), higher BMI (P=.009), greater number of chronic diseases (P=.02), greater number of cigarettes smoked per day (P=.02), current depressive and/or anxiety disorders (P=.05), and greater severity of depressive symptoms (P<.001). A high overall daily activity level was associated with work/school days (P=.02) and summer (reference: winter; P=.03) (Table 2). Earlier morning activity was associated with older age (P=.02), having a partner (P=.009), work/school days (P<.001), and autumn and spring (reference: winter; P=.02 and P<.001, respectively) (Table 2). Monophasic activity was associated with older age (P=.005; Table 2). Biphasic activity was associated with work/school days (P<.001) and summer (reference: winter; P<.001) (Table 2). Earlier biphasic activity was associated with older age (P=.005), work/school days (P<.001), and spring and summer (reference: winter; P<.001 and P=.005, respectively) (Table 2).

Age, work/school status, and season were significantly associated with motor activity in the FoSR analyses (Figure 2; P<.05). Older age was especially related to lower activity in the late afternoon (around 6 PM). Working or going to school was associated with a higher activity level, with a dip in activity level around noon, compared with not working/going to school (Figure 2). Compared with individuals assessed in winter, those assessed in summer, spring, and autumn had higher activity levels in the morning and those assessed in the summer had higher activity levels late in the afternoon (Figure 2). Average daily motor activity curves by age, work/school status, and season are shown in Figure 3.

Table 2. Multivariable associations between daily motor activity patterns and sociodemographics, health and lifestyle factors, actigraphy assessment characteristics, and psychopathology (N=359).[a]

Variables	Overall daily activity intensity			Early vs late morning activity			Biphasic vs monophasic activity			Early vs late biphasic activity		
	β	SE	*P*	β	SE	*P*	β	SE	*P*	B	SE	*P*
Sociodemographics												
Age	–.15	0.04	<.001	.08	0.03	.02	–.10	0.03	.005	.10	0.03	.005
Sex												
Female	Ref[b]			Ref			Ref			Ref		
Male	–.05	0.08	.65	–.02	0.06	.83	–.03	0.06	.79	–.07	0.06	.34
Education level	–.13	0.04	.005	–.01	0.03	.80	.04	0.03	.30	–.03	0.03	.54
Has a partner												
No	Ref			Ref			Ref			Ref		
Yes	–.17	–0.08	.07	.18	–0.06	.009	–.02	–0.05	.83	.05	–0.05	.496
Lifestyle factors												
BMI	–.12	0.04	.009	.05	0.03	.22	–.01	0.02	.83	.01	0.03	.80
Number of chronic diseases	–.11	0.04	.02	–.01	0.03	.82	–.03	0.03	.47	–.05	0.03	.15
Number of cigarettes per day	–.14	0.05	.02	.03	0.03	.55	–.05	0.03	.19	.02	0.03	.69
Number of drinks per day	.08	0.05	.19	–.05	0.04	.40	–.04	0.03	.34	–.01	0.03	.83
Actigraphy assessment characteristics												
Work/school day												
No	Ref			Ref			Ref			Ref		
Yes	.17	0.06	.02	.28	0.05	<.001	.73	0.05	<.001	.19	0.05	<.001
Season												
Winter	Ref			Ref			Ref			Ref		
Autumn	.10	0.09	.39	.19	0.07	.02	.02	0.07	.83	.14	0.06	.07
Spring	.15	0.09	.21	.33	0.07	<.001	.14	0.07	.08	.26	0.07	<.001
Summer	.29	0.11	.03	.18	0.10	.19	.33	0.07	<.001	.24	0.08	.005
Psychopathology												
Depressive and/or anxiety disorders												
No	Ref			Ref			Ref			Ref		
Remitted	–.06	0.08	.60	–.04	0.08	.79	–.05	0.07	.64	.08	0.07	.39
Current	–.24	0.10	.05	.02	0.08	.83	–.001	0.08	.99	–.01	0.07	.95
IDS[c]	–.14	0.04	<.001	–.05	0.04	.34	–.02	0.03	.70	–.04	0.03	.30
Antidepressant use												
No	Ref			Ref			Ref			Ref		
Yes	–.17	0.11	.22	–.12	0.06	.14	–.08	0.07	.38	–.11	0.08	.33

[a]Multivariable generalized estimating equation models were run and each model was adjusted for sociodemographic, health and lifestyle, and sampling factors. Multiple testing correction was applied using false discovery rate estimation.

[b]Ref: reference.

[c]IDS: Inventory of Depressive Symptomatology.

Figure 2. Effect of sociodemographics, health and lifestyle factors, and actigraphy and clinical characteristics on time of activity from multivariable function-on-scalar regression analyses. IDS: Inventory of Depressive Symptomatology. *P<.05; **P<.01.

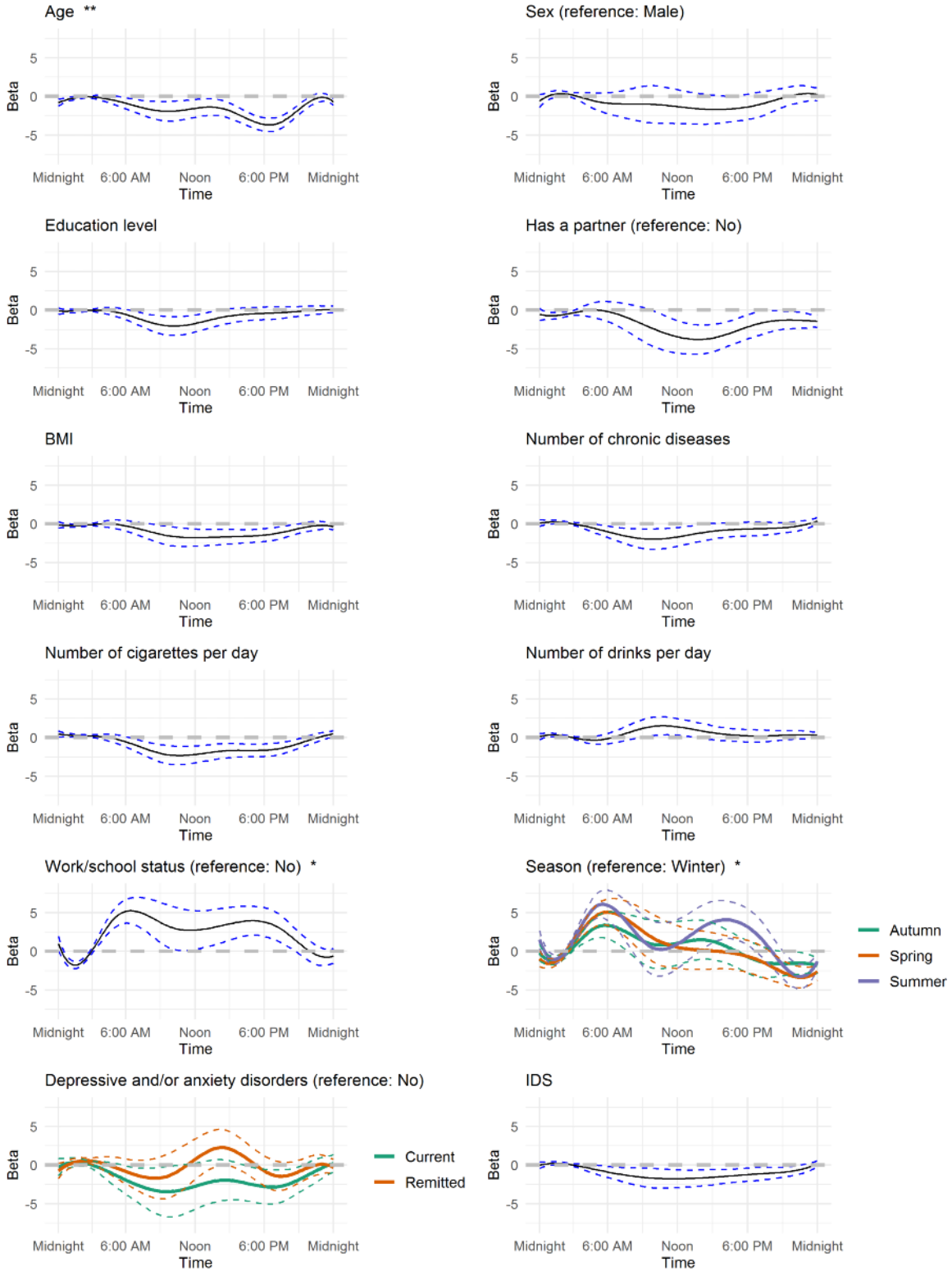
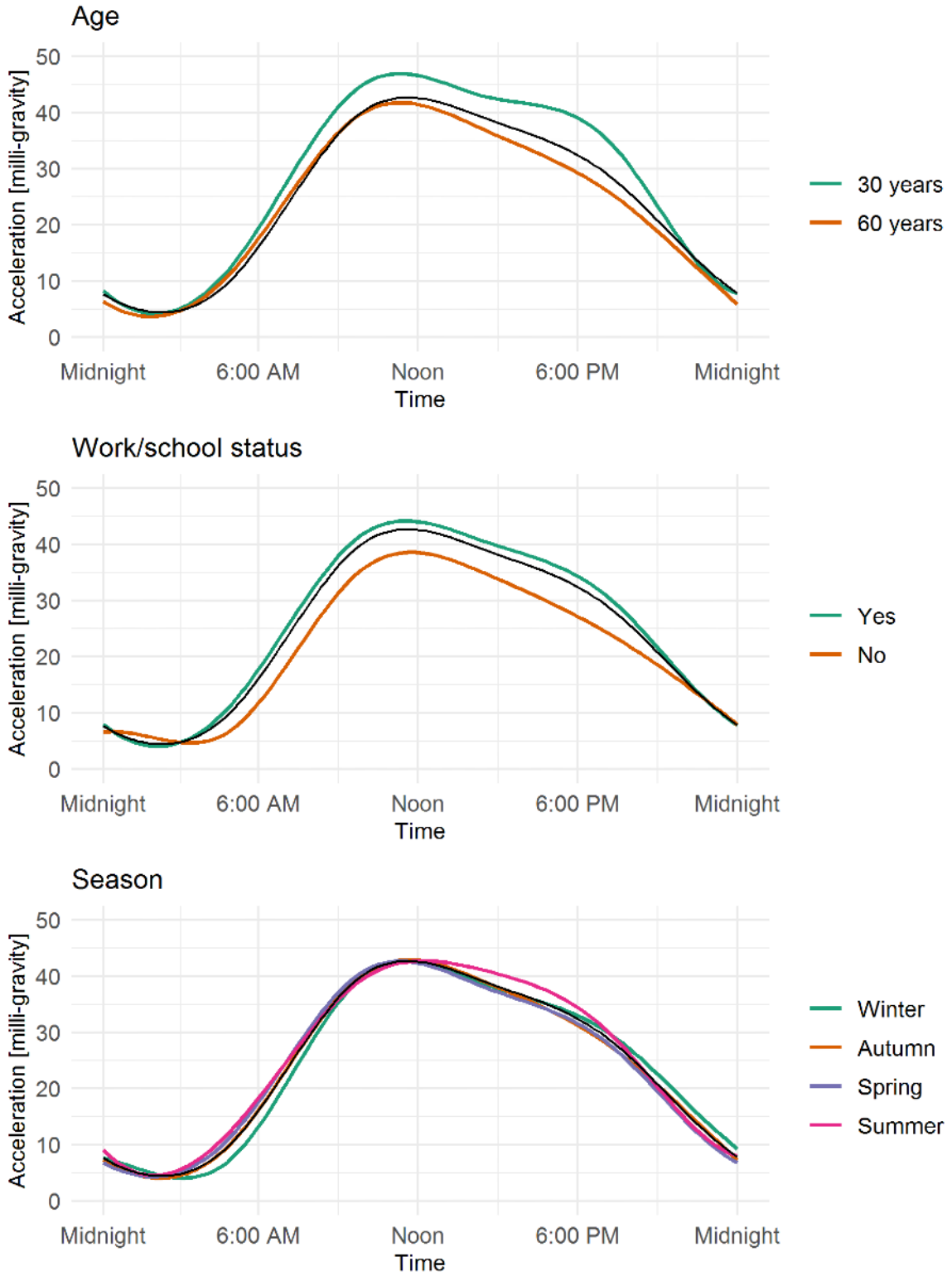


Figure 3. Average daily motor activity curves by age, work/school status, and season.



Discussion

Principal Findings

This is the first study to examine the associations between patterns of daily motor activity and sociodemographic, health and lifestyle, and sampling factors and psychiatric clinical characteristics in individuals with and without depressive and anxiety disorders using FDA. Patterns of daily motor activity extracted with fPCA seem to reflect commonly studied circadian activity rhythm features such as daily activity level and time-of-day preference for morningness or eveningness [36]. The presence and severity of depressive and anxiety disorders were associated with a low overall activity pattern but had no impact on time of activity. Sociodemographic, health and lifestyle, and sampling factors were independently associated with daily motor activity patterns. FoSR analyses indicated that

age, work/school status, and season of assessment were associated with time of activity.

Comparison With Previous Literature

The majority of the variability in the data (77.4%) was explained by four principal components that reflect the complexity of activity patterns. In line with previous studies employing fPCA, the first two components were commonly studied actigraphy features related to circadian rhythms: overall level of physical activity [18,19,33] and time-of-day preference for morningness or eveningness [33]. Interpretation of the third and fourth components was less clear. We found a monophasic versus biphasic activity pattern, as previously reported in other studies [19,33], where it was suggested that a biphasic rhythm may reflect napping behavior. This monophasic pattern could also be related to the temporary drop in alertness and performance that often occurs during the early afternoon, referred to as the postlunch dip, which reflects the 12-hour harmonic of the circadian clock [37]. Therefore, the level of daily activity and chronotype seem to be consistent components across studies applying fPCA, demonstrating the generalizability of the extracted components and confirming that these are important features of daily activity patterns.

Sociodemographic factors, especially age, were associated with daily motor activity patterns with a low activity level. This is in line with the results of the study by Takagi et al [38]. This may be due to the age-related decline in physical activity [39]. Aging causes changes in the organism that lead to a gradual loss of function, frailty, disease, and disability [40], and therefore result in decreased physical activity and physical functioning. The sleep-wake cycle also appears to change in the aging process. Our findings showing the association between earlier rhythms and increasing age are consistent with previous research that shows that aging is associated with advanced sleep timing (as reviewed by Duffy et al [41]) and a preference for morningness (as reviewed by Hood et al [42]). The circadian phase of melatonin has also been reported to become earlier with age, as has the timing of the cortisol rhythm [41]. The suprachiasmatic nucleus (SCN), which represents the biological clock of the brain, shows functional changes with age [43] that may be related to disturbances in circadian rhythms.

Work/school status and season, not surprisingly, were also very important to daily activity patterns and circadian rhythms. Circadian rhythms are controlled centrally by the SCN and influenced externally by behavioral/social cues and by light exposure, as reviewed by Duffy and Czeisler [44]. Our finding that assessments performed on autumn, spring, and summer days showed significantly higher overall levels of activity and higher levels of early morning activity than assessments performed on winter days seems to be consistent with the systematic review of Tucker and Gilliland [45], which showed that level of activity varies with seasonality and is the lowest during winter. Also, other factors related to the season can have an impact on daily activity such as weather conditions, which can also explain different circadian patterns across seasons. Also, on work/school days, there appeared to be higher and more sustained levels of daily activity and earlier morning activity compared with on nonwork/nonschool days. Indeed, it is well known in the literature that modern life habits including night work, shift work [46], jet lag [47], and social jet lag [48] are associated with circadian rhythm disruptions.

We also found that health and lifestyle factors are linked to daily activity patterns. Our results are in line with previous research reporting an association between lower activity levels and higher BMIs in the NIMH Family study [12] and greater numbers of chronic diseases [49]. These results may be suggestive of sedentary behavior, a factor known to relate to weight gain and disabilities. Early morning activity was associated with higher BMIs. This might be because of respiration-related diseases such as apnea, which is known to be more prevalent in persons with a high BMI and can severely disturb sleep [50].

An important aspect of this research was to study the association between timing of daily motor activity and psychopathology. However, we found no association with the timing of the activity. Instead, our functional data-driven models showed similar associations as were found in our previous analyses on NESDA data [7,8], indicating that current depressive and/or anxiety disorders and more severe symptoms were associated with lower physical activity levels but not with a preference for eveningness. These results suggest that the use of daily indices of motor activity may be sufficient when studying the association with psychopathology. On the other hand, we have only evaluated group-level differences; studying differences at the individual level may be important to explore. For instance, by using an idiographic approach (ie, study associations that differ between time points or between individuals), it may be possible to study the dynamics between daily motor activity and depression and/or anxiety [51] and help identify patients in whom activity is strongly predictive of mood. Collecting more empirical data in clinical practice is necessary to establish whether this is a promising approach.

Strengths of the Study

An important strength of this paper is the use of FDA, which is a useful statistical tool for data exploration and visualization. By providing a graphical representation of motor activity and circadian rhythms, FDA can help to identify specific patterns. This could help to generate new hypotheses that could, in turn, contribute to an improvement of the treatment of circadian disturbances. FDA could also be used to predict future outcomes of treatment. For instance, the study of Zeitzer et al [19] showed that low daytime activity and a late afternoon peak extracted with fPCA are predictive of higher mortality rates in community-dwelling older men, although it remains to be investigated whether fPCA components add to the prediction over traditional actigraphy measures. Daily curves of motor activity could potentially also be used in predictive models to pick up early signs of recovery or nonresponse and, if predictive, could inform clinicians in monitoring treatment response or treatment planning.

Limitations

This study was limited by several factors. The data were observational and cross-sectional, and thus the associations cannot be inferred to be causal. As the subjects in this study

were individuals participating in the fifth wave of a prospective cohort, there may be a selection bias toward highly motivated individuals. Future studies may investigate whether our results may be replicated in a wider population with and without depression and anxiety. Although actigraphy provides only an indirect assessment of circadian rhythm, it has the advantage of continuously monitoring over a relatively long period of time. Not all functional components were easily interpretable. While the first two components are (possibly) indicative of the overall level of activity and the time-of-day preference for morningness or eveningness, the third and fourth components need to be replicated in order to provide a better interpretation and greater validity.

Conclusions

Our study showed that features of daily motor activity extracted with fPCA reflect commonly studied factors such as the daily activity level and the time-of-day preference for morningness or eveningness. Age, work/school status, and season were the variables most strongly associated with patterns of daily activity and had time-varying associations with daily motor activity. The presence and severity of depressive and anxiety disorders were associated with a lower overall activity level but not with the timing of activity. Other than psychopathology, sociodemographic, health and lifestyle, and sampling factors were independently associated with a low overall activity pattern.

Acknowledgments

This work was supported by Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115902. FL has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° PCIG12-GA-2012-334065. The infrastructure for the NESDA study (www.nesda.nl) is funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development (ZonMw, grant number 10-000-1002) and financial contributions by participating universities and mental health care organizations (VU University Medical Center, GGZ inGeest, Leiden University Medical Center, Leiden University, GGZ Rivierduinen, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, and Rob Giel Onderzoekscentrum).

According to European law (GDPR), data containing potentially identifying or sensitive patient information are restricted; our data involving clinical participants are not freely available in a public repository. However, data are available upon request via the NESDA Data Access Committee (nesda@ggzingeest.nl).

This research was supported by the Intramural Research Program of the National Institute of Mental Health through grant ZIA MH002954-02 [Motor Activity Research Consortium for Health (mMARCH)].

Authors' Contributions

SD, FL, and BWJHP formulated the research questions. SD performed the data cleaning and the statistical analyses, interpreted the results, wrote the manuscript, and incorporated feedback from all coauthors. FL contributed to the data cleaning, reviewed and provided feedback on all drafts of the manuscript, and critically interpreted the results. BWJHP reviewed and provided feedback on all drafts of the manuscript and critically interpreted the results. HS provided feedback on the statistical analyses, contributed to the interpretation of results, and revised the manuscript. HR, KRM, VZ, NA, AMvH, and RAS contributed to the interpretation of results and revised the manuscript critically for important intellectual content. All authors approved of the final version of the paper.

Conflicts of Interest

None declared.

References

1. Martin JL, Hakim AD. Wrist actigraphy. Chest 2011 Jun;139(6):1514-1527 [FREE Full text] [doi: 10.1378/chest.10-1872] [Medline: 21652563]
2. Ancoli-Israel S, Martin JL, Blackwell T, Buenaver L, Liu L, Meltzer LJ, et al. The SBSM Guide to Actigraphy Monitoring: Clinical and Research Applications. Behav Sleep Med 2015;13 Suppl 1:S4-S38. [doi: 10.1080/15402002.2015.1046356] [Medline: 26273913]
3. Matthews CE, Hagströmer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S68-S76 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399e5b] [Medline: 22157777]
4. Lyall LM, Wyse CA, Graham N, Ferguson A, Lyall DM, Cullen B, et al. Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank. Lancet Psychiatry 2018 Jun;5(6):507-514. [doi: 10.1016/S2215-0366(18)30139-1] [Medline: 29776774]
5. Abbott SM, Zee PC. Circadian Rhythms: Implications for Health and Disease. Neurol Clin 2019 Aug;37(3):601-613. [doi: 10.1016/j.ncl.2019.04.004] [Medline: 31256792]

6. Burton C, McKinstry B, Szentagotai Tătar A, Serrano-Blanco A, Pagliari C, Wolters M. Activity monitoring in patients with depression: a systematic review. J Affect Disord 2013 Feb 15;145(1):21-28. [doi: 10.1016/j.jad.2012.07.001] [Medline: 22868056]
7. Difrancesco S, Lamers F, Riese H, Merikangas KR, Beekman ATF, van Hemert AM, et al. Sleep, circadian rhythm, and physical activity patterns in depressive and anxiety disorders: A 2-week ambulatory assessment study. Depress Anxiety 2019 Oct;36(10):975-986 [FREE Full text] [doi: 10.1002/da.22949] [Medline: 31348850]
8. Minaeva O, Booij SH, Lamers F, Antypa N, Schoevers RA, Wichers M, et al. Level and timing of physical activity during normal daily life in depressed and non-depressed individuals. Transl Psychiatry 2020 Jul 30;10(1):259 [FREE Full text] [doi: 10.1038/s41398-020-00952-w] [Medline: 32732880]
9. Banihashemi N, Robillard R, Yang J, Carpenter JS, Hermens DF, Naismith SL, et al. Quantifying the effect of body mass index, age, and depression severity on 24-h activity patterns in persons with a lifetime history of affective disorders. BMC Psychiatry 2016 Sep 09;16(1):317 [FREE Full text] [doi: 10.1186/s12888-016-1023-2] [Medline: 27612556]
10. Mitchell JA, Quante M, Godbole S, James P, Hipp JA, Marinac CR, et al. Variation in actigraphy-estimated rest-activity patterns by demographic factors. Chronobiol Int 2017;34(8):1042-1056 [FREE Full text] [doi: 10.1080/07420528.2017.1337032] [Medline: 28650674]
11. Cespedes Feliciano EM, Quante M, Weng J, Mitchell JA, James P, Marinac CR, et al. Actigraphy-Derived Daily Rest-Activity Patterns and Body Mass Index in Community-Dwelling Adults. Sleep 2017 Dec 01;40(12):zsx168 [FREE Full text] [doi: 10.1093/sleep/zsx168] [Medline: 29029250]
12. Shou H, Cui L, Hickie I, Lameira D, Lamers F, Zhang J, et al. Dysregulation of objectively assessed 24-hour motor activity patterns as a potential marker for bipolar I disorder: results of a community-based family study. Transl Psychiatry 2017 Aug 22;7(8):e1211 [FREE Full text] [doi: 10.1038/tp.2017.136] [Medline: 28892068]
13. Stothard ER, McHill AW, Depner CM, Birks BR, Moehlman TM, Ritchie HK, et al. Circadian Entrainment to the Natural Light-Dark Cycle across Seasons and the Weekend. Curr Biol 2017 Feb 20;27(4):508-513 [FREE Full text] [doi: 10.1016/j.cub.2016.12.041] [Medline: 28162893]
14. van Maanen A, Meijer AM, van der Heijden KB, Oort FJ. The effects of light therapy on sleep problems: A systematic review and meta-analysis. Sleep Med Rev 2016 Oct;29:52-62. [doi: 10.1016/j.smrv.2015.08.009] [Medline: 26606319]
15. Penders TM, Stanciu CN, Schoemann AM, Ninan PT, Bloch R, Saeed SA. Bright Light Therapy as Augmentation of Pharmacotherapy for Treatment of Depression. Prim. Care Companion CNS Disord 2016 Oct 20;18(5). [doi: 10.4088/pcc.15r01906] [Medline: 27835725]
16. Klerman EB, Wang W, Phillips AJK, Bianchi MT. Statistics for Sleep and Biological Rhythms Research. J Biol Rhythms 2017 Feb;32(1):18-25 [FREE Full text] [doi: 10.1177/0748730416670051] [Medline: 27836939]
17. Harezlak J, Ruppert D, Wand M. Semiparametric Regression with R. Switzerland: Springer Nature; 2018.
18. Zeitzer JM, David R, Friedman L, Mulin E, Garcia R, Wang J, et al. Phenotyping apathy in individuals with Alzheimer disease using functional principal component analysis. Am J Geriatr Psychiatry 2013 Apr;21(4):391-397 [FREE Full text] [doi: 10.1016/j.jagp.2012.12.012] [Medline: 23498386]
19. Zeitzer JM, Blackwell T, Hoffman AR, Cummings S, Ancoli-Israel S, Stone K, Osteoporotic Fractures in Men (MrOS) Study Research Group. Daily Patterns of Accelerometer Activity Predict Changes in Sleep, Cognition, and Mortality in Older Men. J Gerontol A Biol Sci Med Sci 2018 Apr 17;73(5):682-687 [FREE Full text] [doi: 10.1093/gerona/glw250] [Medline: 28158467]
20. Goldsmith J, Zipunnikov V, Schrack J. Generalized multilevel function-on-scalar regression and principal component analysis. Biometrics 2015 Jun 25;71(2):344-353 [FREE Full text] [doi: 10.1111/biom.12278] [Medline: 25620473]
21. Goldsmith J, Liu X, Jacobson JS, Rundle A. New Insights into Activity Patterns in Children, Found Using Functional Data Analyses. Med Sci Sports Exerc 2016 Sep;48(9):1723-1729 [FREE Full text] [doi: 10.1249/MSS.0000000000000968] [Medline: 27183122]
22. Scott J, Murray G, Henry C, Morken G, Scott E, Angst J, et al. Activation in Bipolar Disorders: A Systematic Review. JAMA Psychiatry 2017 Feb 01;74(2):189-196. [doi: 10.1001/jamapsychiatry.2016.3459] [Medline: 28002572]
23. 2018. The Motor Activity Research Consortium for Health (mMARCH). URL: http://mmarch.org/ [accessed 2019-10-07]
24. Penninx BWJH, Beekman ATF, Smit JH, Zitman FG, Nolen WA, Spinhoven P, NESDA Research Consortium. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int J Methods Psychiatr Res 2008;17(3):121-140 [FREE Full text] [doi: 10.1002/mpr.256] [Medline: 18763692]
25. Schoevers RA, van Borkulo CD, Lamers F, Servaas MN, Bastiaansen JA, Beekman ATF, et al. Affect fluctuations examined with ecological momentary assessment in patients with current or remitted depression and anxiety disorders. Psychol Med 2020 Apr 01:1-10. [doi: 10.1017/S0033291720000689] [Medline: 32234092]
26. da Silva IC, van Hees VT, Ramires VV, Knuth AG, Bielemann RM, Ekelund U, et al. Physical activity levels in three Brazilian birth cohorts as assessed with raw triaxial wrist accelerometry. Int J Epidemiol 2014 Dec;43(6):1959-1968 [FREE Full text] [doi: 10.1093/ije/dyu203] [Medline: 25361583]
27. Wittchen HU. Reliability and validity studies of the WHO--Composite International Diagnostic Interview (CIDI): a critical review. J Psychiatr Res 1994;28(1):57-84. [doi: 10.1016/0022-3956(94)90036-1] [Medline: 8064641]

28. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med 1996 May;26(3):477-486. [Medline: 8733206]
29. van Hees V. Accelerometer data processing with GGIR. URL: https://cran.r-project.org/web/packages/GGIR/vignettes/GGIR.html [accessed 2017-10-01]
30. van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva ICM, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. J Appl Physiol (1985) 2014 Oct 01;117(7):738-744 [FREE Full text] [doi: 10.1152/japplphysiol.00421.2014] [Medline: 25103964]
31. van Hees VT, Gorzelniak L, Dean León EC, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One 2013;8(4):e61691 [FREE Full text] [doi: 10.1371/journal.pone.0061691] [Medline: 23626718]
32. Ramsay J, Graves S, Hooker G. fda: Functional Data Analysis. URL: https://cran.r-project.org/web/packages/fda/index.html [accessed 2018-04-27]
33. Gershon A, Ram N, Johnson SL, Harvey AG, Zeitzer JM. Daily Actigraphy Profiles Distinguish Depressive and Interepisode States in Bipolar Disorder. Clin Psychol Sci 2016 Jul;4(4):641-650 [FREE Full text] [doi: 10.1177/2167702615604613] [Medline: 27642544]
34. Noble WS. How does multiple testing correction work? Nat Biotechnol 2009 Dec;27(12):1135-1137 [FREE Full text] [doi: 10.1038/nbt1209-1135] [Medline: 20010596]
35. Goldsmith J. Function-on-Scalar Regression. URL: http://jeffgoldsmith.com/IWAFDA/shortcourse_fosr.html [accessed 2019-05-24]
36. Kivelä L, Papadopoulos MR, Antypa N. Chronotype and Psychiatric Disorders. Curr Sleep Med Rep 2018;4(2):94-103 [FREE Full text] [doi: 10.1007/s40675-018-0113-8] [Medline: 29888167]
37. Monk TH. The post-lunch dip in performance. Clin Sports Med 2005 Apr;24(2):e15-23, xi. [doi: 10.1016/j.csm.2004.12.002] [Medline: 15892914]
38. Takagi D, Nishida Y, Fujita D. Age-associated changes in the level of physical activity in elderly adults. J Phys Ther Sci 2015 Dec;27(12):3685-3687 [FREE Full text] [doi: 10.1589/jpts.27.3685] [Medline: 26834332]
39. Sallis JF. Age-related decline in physical activity: a synthesis of human and animal studies. Med Sci Sports Exerc 2000 Sep;32(9):1598-1600. [doi: 10.1097/00005768-200009000-00012] [Medline: 10994911]
40. Paterson DH, Warburton DE. Physical activity and functional limitations in older adults: a systematic review related to Canada's Physical Activity Guidelines. Int J Behav Nutr Phys Act 2010 May 11;7:38 [FREE Full text] [doi: 10.1186/1479-5868-7-38] [Medline: 20459782]
41. Duffy JF, Zitting K, Chinoy ED. Aging and Circadian Rhythms. Sleep Med Clin 2015 Dec;10(4):423-434 [FREE Full text] [doi: 10.1016/j.jsmc.2015.08.002] [Medline: 26568120]
42. Hood S, Amir S. The aging clock: circadian rhythms and later life. J Clin Invest 2017 Feb 01;127(2):437-446 [FREE Full text] [doi: 10.1172/JCI90328] [Medline: 28145903]
43. Peters R. Ageing and the brain. Postgrad Med J 2006 Feb;82(964):84-88 [FREE Full text] [doi: 10.1136/pgmj.2005.036665] [Medline: 16461469]
44. Duffy JF, Czeisler CA. Effect of Light on Human Circadian Physiology. Sleep Med Clin 2009 Jun;4(2):165-177 [FREE Full text] [doi: 10.1016/j.jsmc.2009.01.004] [Medline: 20161220]
45. Tucker P, Gilliland J. The effect of season and weather on physical activity: a systematic review. Public Health 2007 Dec;121(12):909-922. [doi: 10.1016/j.puhe.2007.04.009] [Medline: 17920646]
46. James SM, Honn KA, Gaddameedhi S, Van Dongen HPA. Shift Work: Disrupted Circadian Rhythms and Sleep-Implications for Health and Well-Being. Curr Sleep Med Rep 2017 Jun;3(2):104-112 [FREE Full text] [doi: 10.1007/s40675-017-0071-6] [Medline: 29057204]
47. Vosko AM, Colwell CS, Avidan AY. Jet lag syndrome: circadian organization, pathophysiology, and management strategies. Nat Sci Sleep 2010;2:187-198 [FREE Full text] [doi: 10.2147/NSS.S6683] [Medline: 23616709]
48. Wittmann M, Dinich J, Merrow M, Roenneberg T. Social jetlag: misalignment of biological and social time. Chronobiol Int 2006;23(1-2):497-509. [doi: 10.1080/07420520500545979] [Medline: 16687322]
49. Durstine JL, Gordon B, Wang Z, Luo X. Chronic disease and the link to physical activity. Journal of Sport and Health Science 2013 Mar;2(1):3-11. [doi: 10.1016/j.jshs.2012.07.009]
50. Romero-Corral A, Caples S, Lopez-Jimenez F, Somers V. Interactions Between Obesity and Obstructive Sleep Apnea. Chest 2010 Mar;137(3):711-719. [doi: 10.1378/chest.09-0360] [Medline: 20202954]
51. Rosmalen JGM, Wenting AMG, Roest AM, de Jonge P, Bos EH. Revealing causal heterogeneity using time series analysis of ambulatory assessments: application to the association between depression and physical activity after myocardial infarction. Psychosom Med 2012 May;74(4):377-386. [doi: 10.1097/PSY.0b013e3182545d47] [Medline: 22582335]

Abbreviations

ATC: Anatomical Therapeutic Chemical
CIDI: Composite International Diagnostic Interview

EMA: ecological momentary assessment
FDA: functional data analysis
FoSR: function-on-scalar regression
fPCA: functional principal component analysis
GEE: generalized estimating equation
IDS: Inventory of Depressive Symptomatology
mMARCH: Motor Activity Research Consortium for Health
NESDA: Netherlands Study of Depression and Anxiety
NESDA-EMAA: Ecological Momentary Assessment and Actigraphy substudy of the Netherlands Study of Depression and Anxiety
SCN: suprachiasmatic nucleus

Edited by G Eysenbach, R Kukafka; submitted 27.05.20; peer-reviewed by J Zeitzer, R Walmsley; comments to author 17.06.20; revised version received 05.08.20; accepted 07.12.20; published 17.02.21.

Please cite as:
Difrancesco S, Riese H, Merikangas KR, Shou H, Zipunnikov V, Antypa N, van Hemert AM, Schoevers RA, Penninx BWJH, Lamers F
Sociodemographic, Health and Lifestyle, Sampling, and Mental Health Determinants of 24-Hour Motor Activity Patterns: Observational Study
J Med Internet Res 2021;23(2):e20700
URL: http://www.jmir.org/2021/2/e20700/
doi:10.2196/20700
PMID:33595445

Review

Framework for the Design Engineering and Clinical Implementation and Evaluation of mHealth Apps for Sleep Disturbance: Systematic Review

Melissa Aji[1], BPsych; Christopher Gordon[2,3], PhD; Elizabeth Stratton[1], PhD; Rafael A Calvo[4], PhD; Delwyn Bartlett[1,2], PhD; Ronald Grunstein[2,5], MD, PhD; Nick Glozier[6], MBBS, PhD

[1]Central Clinical School, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[2]CIRUS, Centre for Sleep and Chronobiology, Woolcock Institute of Medical Research, Glebe, Australia

[3]Susan Wakil School of Nursing and Midwifery, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[4]Dyson School of Design Engineering, Imperial College London, London, United Kingdom

[5]Charles Perkins Centre - RPA Clinic, Royal Prince Alfred Hospital, Sydney, Australia

[6]Brain and Mind Center, The University of Sydney, Camperdown, Australia

Corresponding Author:
Nick Glozier, MBBS, PhD
Brain and Mind Center
The University of Sydney
Level 5, Professor Marie Bashir Centre, Missenden Road
Camperdown, 2050
Australia
Phone: 61 29515 1596
Email: nick.glozier@sydney.edu.au

Abstract

Background: Mobile health (mHealth) apps offer a scalable option for treating sleep disturbances at a population level. However, there is a lack of clarity about the development and evaluation of evidence-based mHealth apps.

Objective: The aim of this systematic review was to provide evidence for the design engineering and clinical implementation and evaluation of mHealth apps for sleep disturbance.

Methods: A systematic search of studies published from the inception of databases through February 2020 was conducted using 5 databases (MEDLINE, Embase, Cochrane Library, PsycINFO, and CINAHL).

Results: A total of 6015 papers were identified using the search strategy. After screening, 15 papers were identified that examined the design engineering and clinical implementation and evaluation of 8 different mHealth apps for sleep disturbance. Most of these apps delivered cognitive behavioral therapy for insomnia (CBT-I, n=4) or modified CBT-I (n=2). Half of the apps (n=4) identified adopting user-centered design or multidisciplinary teams in their design approach. Only 3 papers described user and data privacy. End-user acceptability and engagement were the most frequently assessed implementation metrics. Only 1 app had available evidence assessing all 4 implementation metrics (ie, acceptability, engagement, usability, and adherence). Most apps were prototype versions (n=5), with few matured apps. A total of 6 apps had supporting papers that provided a quantitative evaluation of clinical outcomes, but only 1 app had a supporting, adequately powered randomized controlled trial.

Conclusions: This is the first systematic review to synthesize and examine evidence for the design engineering and clinical implementation and evaluation of mHealth apps for sleep disturbance. The minimal number of apps with published evidence for design engineering and clinical implementation and evaluation contrasts starkly with the number of commercial sleep apps available. Moreover, there appears to be no standardization and consistency in the use of best practice design approaches and implementation assessments, along with very few rigorous efficacy evaluations. To facilitate the development of successful and evidence-based apps for sleep disturbance, we developed a high-level framework to guide researchers and app developers in the end-to-end process of app development and evaluation.

(J Med Internet Res 2021;23(2):e24607) doi:10.2196/24607

KEYWORDS

mobile applications; sleep; insomnia; internet-based intervention; mHealth; mobile health; systematic review

Introduction

Sleep disturbance is extremely prevalent and affects 33%-45% of adults [1]. Insomnia is the most common sleep disorder, defined as a chronic and persistent difficulty falling asleep, maintaining sleep, or waking up too early [2]. When left untreated, insomnia significantly increases the risk of adverse health outcomes, including mental health disorders [3,4], cardiovascular disease [5], hypertension [6], and diabetes [7]. As insomnia poses serious risks to mental and physical health, exploring the efficacy of treatment is imperative. Cognitive behavioral therapy for insomnia (CBT-I) is an effective gold standard treatment that has consistently shown moderate to large treatment effects [8-10]. A limitation to the widespread use of CBT-I has been a shortage of adequately trained practitioners to treat the high volume of patients with insomnia.

One potential avenue for addressing these challenges is the use of digital therapy. The widespread adoption of mobile phones and apps can change the delivery of health care. This emerging field, known as mobile health (mHealth), refers to the provision of health care services and practice delivered using mobile technology. mHealth apps provide unique benefits in delivering health information and interventions, given the ubiquity, convenience, and affordability of mobile phones. Over 325,000 mHealth apps are available on app stores, and this number continues to grow rapidly [11]. Research supports the utility of mHealth apps for a range of health issues, including depression, anxiety, schizophrenia, cardiac disease, physical activity, and diabetes [12-16]. The potential of mHealth is particularly significant for sleep disturbances as it presents a promising method for addressing this public health burden.

Health outcomes are not only dependent on the intervention delivered but also on the design engineering process employed. Design engineering combines design thinking, participatory design practices, software engineering methods, software, and quality assurance methods. Without analyzing this process, it is not possible to make inferences about the reasons for failure (eg, low engagement, lack of clinical efficacy) of an app. Alongside, this process involves the incorporation of clinically relevant content that either provides adjunctive or standalone therapy to traditional medical and psychological practice.

Clinical implementation requires consideration of end-user privacy and security, a major concern in mHealth. Recent work has demonstrated that mHealth apps routinely share and commercialize end-user data with third parties with very little transparency [17]. In addition, there are several important regulatory considerations required to protect the public and ensure that mHealth apps meet the minimum requirements of quality. Regulatory bodies for mHealth include HIPAA (Health Insurance Portability and Accountability Act), which is a federal law mandating privacy and security standards, and the FDA (Food and Drug Administration), which evaluates the safety and marketing claims of mHealth apps. In the absence of oversight from these regulatory bodies, it is possible to misuse health care–related data [18].

Despite the potential and ubiquity of mHealth apps, most apps lack evidence for their clinical efficacy among end users [15]. Compounding this problem is a lack of framework to inform and standardize the process and reporting of design, development, and evaluation of mHealth apps [19]. This may lead to clinical inefficacy, lack of medical condition–specific content, poor patient engagement, or even harmful apps [20,21].

Although various research groups have established frameworks for apps for posttraumatic stress disorder (PTSD) [22], bipolar disorder [23], and hypertension [24], there may be differences in the processes used according to the health condition, and there are no existing frameworks for sleep apps. This systematic review aims to assess the extent and nature of the peer-reviewed evidence and proposes a high-level framework for the design engineering and clinical implementation and evaluation of mHealth apps for sleep disturbance.

Methods

Search Strategy

This study uses PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines [25]. A systematic literature search was conducted using the electronic databases MEDLINE, Embase, Cochrane Library, PsycINFO, and CINAHL for relevant papers published from the inception of the databases through February 2020. Keywords related to sleep and mHealth were searched. To increase the coverage of relevant databases, we conducted a manual search of the Journal of Medical Internet Research and the Journal of Internet Interventions as well as the reference list of retrieved publications. See Multimedia Appendix 1 for an example of the search strategy.

Inclusion Criteria

The inclusion criteria consisted of peer-reviewed publications that (1) focused on mHealth apps that aimed to measure, track, or improve sleep; (2) described the design engineering, clinical implementation, or clinical evaluation of the apps; (3) focused on an app solely targeting sleep; and (4) targeted adults aged between 18 years and 60 years. The included studies were published in English.

Exclusion Criteria

Studies were excluded if they (1) focused on a sleep disorder other than insomnia; (2) focused on an app that provided multimodal interventions targeting health aspects other than sleep; (3) described internet, telephone, or text messaging interventions; or (4) were review papers. We excluded review papers that were at the early stages of screening; however, given the difficulty of discerning these papers, most were excluded in the full-text screening.

Screening

After duplicates were removed, a single author (MA) screened all the titles to identify potentially relevant studies. Abstract and title screen and full-text of potentially relevant studies were reviewed independently by 2 authors (MA and ES). Where there were conflicts, discrepancies were discussed, and consensus was reached with the senior author (NG).

Data Extraction and Coding

A data extraction template was constructed by one of the authors (MA) to summarize the following characteristics of the studies: first author name, publication date, country, objective, study design, and sample. Full papers were then imported into NVivo (QSR International Pty Ltd, version 11.4.3, 2017) for detailed data extraction and thematic coding. A coding framework was developed by the authors for each of the following categories: design engineering and clinical implementation and evaluation. Relevant segments of text were coded and extracted into a Microsoft Excel spreadsheet. Risk of bias was not assessed, as this was deemed irrelevant to the aims of the paper.

Data were extracted and coded into the following subgroups:

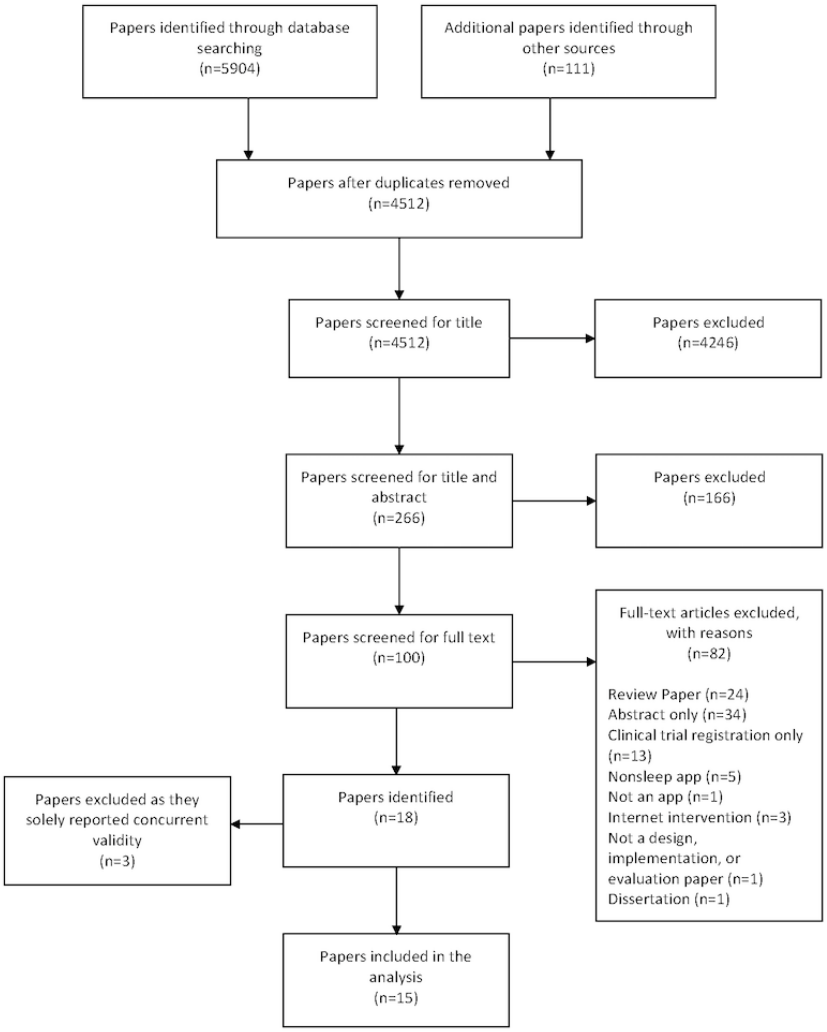
- Design engineering: This stage considers the therapeutic approach, design method, and features and functionalities of the apps. Although we appreciate the design engineering process as an ongoing and iterative process, for the purpose of this paper, we propose that this stage covers the point up until the prototype app is implemented among end users.
- Clinical implementation: This stage refers to the testing of the app among end users and includes an assessment of app maturity, implementation metrics (acceptability, engagement, usability, and adherence), and privacy (ie, data acquisition, use or disclosures of identifiable health data) and regulatory (eg, HIPAA) requirements. To categorize the implementation metrics used by the papers and uniformly compare evidence in the included studies, we developed definitions based on previous literature and adjusted for use with sleep apps [26-29]. Acceptability refers to how intended recipients react to the intervention, for example, interest, user satisfaction, and perceived appropriateness. Engagement refers to the usage metrics of the app. Usability refers to whether the app can be used as intended. Adherence refers to the degree to which the user follows the program and the prescribed recommendations of the therapy.
- Clinical evaluation: This stage refers to the assessment of treatment outcomes among end users.

Results

Search Results

The search strategy identified 6015 papers (Figure 1). Following the removal of duplicates (n=1503), 4512 papers were screened for title. Of those, 4246 were excluded, leaving 266 potentially relevant papers. Following a duplicate independent abstract review, 166 papers were excluded. The full text of 100 papers was obtained and independently reviewed by 2 authors (MA and ES). Following this, 82 papers were excluded, primarily for being review papers (n=24) or abstracts only (n=34), leaving 18 full-text papers to be included in this systematic review. A further 3 papers were excluded that reported evidence solely for concurrent validity *(ie, comparison of app sleep tracking against objective measurements of sleep)* and provided no further information about design engineering and clinical implementation and evaluation.

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



Study Characteristics

A summary of the study characteristics is presented in Table 1. There was a total of 15 papers. These papers assessed the design, implementation, and evaluation of 8 individual apps.

Furthermore, there were multiple studies (n=27) included within these 15 papers: 7 pre-post studies, 5 cross-sectional studies, 5 mixed methods studies, 5 concurrent validity studies, 2 randomized controlled trials (RCTs), and 3 descriptive studies (Table 1). Table 1 presents the papers and corresponding studies numbered under the column *Number of studies and their objectives*. Subjects were veterans or military service members (n=5) [30-34], healthy participants (n=6) [35-40], individuals with insomnia or sleep disturbances (n=4) [41-44], veterans affairs CBT-I clinicians (n=2) [45,46], and nurses with insomnia (n=1) [47]. The sample size varied among studies, ranging from 2 to 176. For longitudinal studies with over 10 participants, the mean study attrition rate was 25% (range 16%-39%).

Table 1. Overview of the included studies.

Study	Country of origin	App name	Sample	Number of studies and their objectives	Study design
Aji et al (2019) [42]	Australia	SleepFix	Participants with insomnia or sleep disturbance	• #1-3 - Assess the needs and preferences of those with poor sleep and insomnia	• #1- Cross-sectional: survey • #2 - Cross-sectional: focus-groups • #3 - Cross-sectional: analysis of app reviews
Aji et al (2020) [43]	Australia	SleepFix	Participants with insomnia	• #1 - Examine the engagement, acceptability, and preliminary efficacy among end users of SleepFix	• #1 - Mixed methods: pre-post and interview
Babson et al (2015) [34]	United States	CBT-I[a] Coach	Veterans with sleep disturbance and cannabis use disorder	• #1 - Examine the acceptability, usability, engagement, and preliminary efficacy among end users of CBT-I Coach	• #1 - Pre-post
Baron et al (2018) [39]	United States	Sleep Bunny	Healthy participants with sleep duration <7 hours	• #1 - Examine the acceptability and adherence with the prototype version among end users of Sleep Bunny • #2 - Examine the acceptability and adherence with the final version among end users of Sleep Bunny	• #1 - Mixed methods: pre-post and interview • #2 - Mixed methods: pre-post and interview
Bauer et al (2012) [40]	United States	ShutEye	Health participants	• #1 - Describe ShutEye app • #2 - Examine the acceptability and adherence among end users of ShutEye	• #1 - Descriptive report • #2 - Mixed methods: pre-post and interview
Horsch et al (2017) [44]	The Netherlands	Sleepcare	Participants with mild insomnia	• #1 - Assess the efficacy of Sleepcare	• #1 - RCT[b]
Kang et al (2017) [41]	Korea	Unnamed	Participants with insomnia	• #1 - Evaluate the efficacy of simplified group CBT-I delivered using a mobile app	• #1 - Pre-post
Koffel et al (2016) [33]	United States	CBT-I Coach	Referrals from veterans affair medical center	• #1 - Examine the acceptability, adherence, engagement, and preliminary efficacy among end users of CBT-I Coach	• #1 - RCT
Kuhn et al (2016) [45]	United States	CBT-I Coach	Veterans affairs CBT-I clinicians	• #1 - Describe CBT-I Coach app • #2 - Assess acceptability of the app with clinicians (before and after usage)	• #1 - Descriptive report • #2 - Cross-sectional surveys
Miller et al (2019) [46]	United States	CBT-I Coach	Veterans affairs CBT-I clinicians	• #1 - Assess acceptability of the app with clinicians (after app usage)	• #1 - Cross-sectional survey
Omeogu et al (2020) [47]	United States	CBT-I Coach	Female veterans administration nurses with insomnia	• #1 - Examine the efficacy of CBT-I Coach	• #1 - Pre-post
Pulantara et al (2018) [31]	United States	iREST[c]	Active duty service members and veterans	• #1 - Describe iREST • #2 - Examine usability among end users of iREST • #3 - Assess the concurrent validity of iREST against a wearable	• #1 - Descriptive report • #2- Pre-post • #3 - Comparison of app sleep diary against Fitbit
Pulantara et al (2018) [30]	United States	iREST	Military service members and veterans	• #1 - Evaluate the efficacy of just-in-time adaptive intervention	• #1 - Pre-post

Study	Country of origin	App name	Sample	Number of studies and their objectives	Study design
Reilly et al (2019) [32]	United States	CBT-I Coach	Veterans with insomnia	• #1 - Evaluate the efficacy of CBT-I Coach	• #1 - Pre-post
Shirazi et al (2013) [35]	Germany	Somnometer	University students	• #1 - Assess the acceptability, engagement, and usability of the Somnometer app • #2 - Assess the concurrent validity of Somnometer • #3 - Collect user's app usage patterns in-the-wild	• #1 - Mixed methods: pre-post and interview • #2 - Comparison of app against actigraphy • #3 - Pre-post

[a]CBT-I: cognitive behavioral therapy for insomnia.

[b]RCT: randomized controlled trial.

[c]iREST: Interactive Resilience Enhancing Sleep Tactics.

Design Engineering

Table 2 summarizes the 8 individual apps identified in this review: CBT-I Coach [33,34,45,46], Somnometer [35], Interactive Resilience Enhancing Sleep Tactics (iREST) [30,31], ShutEye [40], Sleepcare [44], Sleep Bunny [39], SleepFix [42,43], and 1 unnamed app [41].

Table 2. Design characteristics of the included mobile apps.

App name	Therapeutic approach	Design method	Level of automation
CBT-I[a] Coach	CBT-I including sleep restriction therapy, stimulus control, psychoeducation, cognitive restructuring, and relapse prevention	•Multidisciplinary team, including experts in insomnia, VA[b]-trained CBT-I clinicians, clinical intervention development, technology, and implementation science • Clinician involvement in design	Fully automated. However, it is not designed to replace clinician-delivered CBT-I; nonetheless, the app can be used as an educational resource
iREST[c]	Traditional behavioral insomnia workflows, brief behavioral therapy for insomnia, a military study, and previous implementation of the JITAI[d] platform	• Clinician involvement in design • Developed based on design principles: iterative and incremental development software model	High clinician input regarding treatment, information, recommendations, and messaging
ShutEye	Mobile, peripheral display to provide real-time sleep hygiene recommendations, relevant to the users' set bed and wake time goals	N/A[e]	Fully automated
Sleepcare	CBT-I including sleep restriction, relaxation, and sleep hygiene	• End-user involvement • Developed based on persuasive strategies: talk-and-tool design principle	Fully automated
SleepFix	Sleep restriction therapy and stimulus control	• End-user involvement • Multidisciplinary team	Fully automated
Sleep Bunny	CBT-I and motivational interviewing and a telephone coaching manual developed from a web-based depression intervention	N/A	Fully automated
Somnometer	Social alarm clock app. Users rate their sleep quality and specify their sleep status, which they can share with their social network. Sleep duration is estimated based on the user's phone interaction with app	N/A	Fully automated
Unnamed	CBT-I including sleep restriction therapy and stimulus control therapy	N/A	Clinician input required

[a]CBT-I: cognitive behavioral therapy for insomnia.

[b]VA: veterans affairs.

[c]iREST: Interactive Resilience Enhancing Sleep Tactics.

[d]JITAI: just-in-time adaptive intervention.

[e]N/A: not applicable.

Therapeutic Approach

A total of 4 apps were identified to deliver a CBT-I intervention, 1 app used behavioral therapy for insomnia (BTi), 1 app delivered sleep restriction therapy (SRT), 1 app was a social alarm clock, and 1 app was a wallpaper display to promote healthy sleep behaviors. The average length of intervention (ie, use of app in the paper) was 5 weeks (mean 4.5. SD 1.5; median 4.5, range 3-6.5 weeks).

Design Approach

The apps were developed using various design approaches, including multidisciplinary (n=2), clinician involvement (n=2), user-centered (n=1), talk-and-tool design principle (n=1), and iterative and incremental development software model (n=1; Table 2). A total of 4 apps did not specify their design approach in their corresponding papers.

Features and Functionality

Table 3 presents the features and functionality of the apps. All apps provided personalized sleep feedback (8/8, 100%). Most apps included psychoeducation or sleep hygiene (7/8, 88%), a sleep diary (6/8, 75%), reminders (5/8, 63%), and visualization features (5/8, 63%). In total, 5 of the 8 apps were fully automated and 3 required clinician input. Papers describing 6 apps provided end-user feedback regarding design [31,33,35,39,40,43]. End-user feedback was generally positive in nature, although some papers reported some negative feedback in the design process (4/6, 67%) [31,39,40,43].

Table 3. App features and functionalities included in the apps (N=8).

Feature	Apps in which the features were included, n (%)	Apps
Personalized sleep feedback	8 (100)	CBT-I[a] Coach, iREST[b], ShutEye Sleepcare, SleepFix, Sleep Bunny, Somnometer, and unnamed app
Psychoeducation or sleep hygiene	7 (88)	CBT-I Coach, iREST, ShutEye, Sleepcare, SleepFix, Sleep Bunny, and unnamed app
Sleep diary	6 (75)	CBT-I Coach, iREST, Sleepcare, SleepFix, Sleep Bunny, and unnamed app
Reminders	5 (63)	CBT-I Coach, iREST, Sleepcare, Sleep Bunny, and SleepFix
Visualization	5 (63)	CBT-I Coach, iREST, SleepFix, Sleep Bunny, and Somnometer
Wearable synchronization	4 (50)	iREST, SleepFix, Sleep Bunny, and unnamed app
Relaxation	3 (38)	CBT-I Coach, Sleepcare, and unnamed app
Alarm	2 (25)	CBT-I Coach, and Somnometer
Clinician portal	2 (25)	iREST and unnamed app
Social features	2 (25)	SleepFix and Somnometer
Messaging with clinician	1 (13)	iREST

[a]CBT-I: cognitive behavioral therapy for insomnia.

[b]iREST: Interactive Resilience Enhancing Sleep Tactics.

Clinical Implementation

Implementation Metrics

All 8 apps included in this review had at least one paper assessing an implementation metric. A total of 10 out of the 15 papers reported on one or more of 4 implementation metrics: acceptability, usability, adherence, and engagement (Table 4).

Table 4. Attributes related to implementation in included papers (N=15).

Attribute	Papers assessing attribute, n (%)	Described apps	Measures
Acceptability by end user	6 (40)	CBT-I[a] Coach, ShutEye, Sleep Bunny, SleepFix, and unnamed app	Interview [33,35,40,43], survey [35,39,41], and MAUM[b] questionnaire [34]
Acceptability by clinician	2 (13)	CBT-I Coach	17-item measure of clinicians' retrospective perceptions of app [45] and 35-item measure of clinicians' retrospective perceptions of app [46]
Engagement	6 (40)	CBT-I Coach, iREST[c], SleepFix, Somnometer, and unnamed app	Self-reported number of sleep diaries [33,41]; MAUM questionnaire measuring frequency of app use and length of use per session [34]; number of sleep diary entries, mood logger entries, and game sessions measured by the app [43]; number of scheduled alarms, posts shared on Facebook, and sessions measured by the app [35]; and number of days the app was used, completion time for each log, number of devices with app accessing server each day measured by the app [31]
Usability	4 (27)	CBT-I Coach, iREST, ShutEye, SleepFix, and Somnometer	System Usability Scale [31,35,43], modified Telerehabilitation Usability Questionnaire [31], MAUM questionnaire [34], qualitative survey [31], and qualitative interview [40]
Adherence	3 (20)	CBT-I Coach, ShutEye, Sleep Bunny, and Sleepcare	Average number of days and time spent on homework [33], patient adherence form (completed by clinicians) [33], number of relaxation exercises performed [44], deviation between real and agreed-upon time in bed [44], number of coaching sessions completed [39], and qualitative interview [40]

[a]CBT-I: cognitive behavioral therapy for insomnia.

[b]MAUM: Mobile App Use Measure.

[c]iREST: Interactive Resilience Enhancing Sleep Tactics.

Acceptability among end users and engagement were the most frequently assessed implementation metrics. Most papers measured acceptability through an interview or survey with end users and engagement using app-measured or self-reported app usage metrics. Usability was measured by less than half of the papers, with most using the System Usability Scale. Most apps had evidence available for 1 to 2 implementation metrics, and only 3 out of 8 apps (CBT-I Coach, ShutEye, and SleepFix) had papers reporting more than 2 implementation metrics.

App Maturity

The maturity levels of the apps ranged on a continuum of 3 levels: prototype, matured, and released [48]. A prototype refers to a minimally viable product of the app with functionality that users can test. A matured version refers to an app that has undergone user testing and has been redesigned. A released version refers to an app that is available for download. Most apps were prototypes (n=5), and 1 app was a prototype-to-mature version (Table 5). Only 1 app was matured to release (CBT-I Coach). Furthermore, 1 study had no information regarding the maturity level of the app described. Only CBT-I Coach was available from the Google Play or Apple App Store at the time of writing the primary paper.

Table 5. Maturity of the included apps.

Stage of maturity	App
Prototype	iREST[a], ShutEye, Sleep Bunny, SleepFix, and Somnometer
Prototype to matured	Sleepcare
Matured	—[b]
Matured to released	CBT-I[c] Coach
Released	—
No information	Unnamed [41]

[a]iREST: Interactive Resilience Enhancing Sleep Tactics.

[b]No apps met this maturity level.

[c]CBT-I: cognitive behavioral therapy for insomnia.

Only 3 apps considered the privacy of data collected [35,40,45]. In total, 3 out of the 4 apps that enabled wearable

synchronization reported data relating to its use adjunct to the app [31,41,43]. No studies have considered the regulatory requirements of the apps.

Clinical Evaluation

In total, 6 of the 8 apps had a quantitative evaluation of treatment outcomes. Of the 11 studies that evaluated clinical effectiveness, the most frequently used sleep outcomes were self-reported sleep questionnaires, followed by app sleep diary measures and actigraphy (objective). In total, 8 of the 11 papers used the Insomnia Severity Index, the most widely used insomnia treatment outcome questionnaire to evaluate insomnia symptom severity [49] (Table 6). Only 2 papers focused solely on evaluating the effectiveness of the app [44,47].

Table 6. Clinical outcome measures used in included papers.

Measures	Study
Sleep	
Self-report questionnaires	
Insomnia Severity Index	[30-33,41,43,44,47]
Pittsburgh Sleep Quality Index	[30,32,34,41,43,44]
Epworth Sleepiness Scale	[30,39,40,43]
Dysfunctional Beliefs and Attitudes about Sleep-16	[41,44]
Functional Outcomes of Sleep Questionnaire	[32]
App sleep diary measures	
Time in bed	[44]
Total sleep time	[44]
Sleep efficiency (percentage of time spent asleep while in bed)	[41,43,44]
Wake after sleep onset (wake time after initial sleep onset)	[43]
Sleep onset latency (time taken to fall asleep)	[41,43]
Actigraphy (objective)	
Sleep efficiency	[39,41]
Total sleep time	[39,41]
Stages of sleep	[41]
Number of awakenings	[32]
Fitbit	
Wake after sleep onset	[43]
Sleep efficiency	[43]
Psychological measures	
Depression and anxiety	
Hospital Anxiety and Depression Scale	[43,44]
Center for Epidemiologic Studies Depression Scale	[44]
PHQ[a]-8, PHQ-9, or PHQ-15	[30,32]
Generalized Anxiety Disorder Scale-7	[30]
Beck Depression Inventory	[41]
Beck Anxiety Inventory	[41]
PTSD[b]	
PTSD Checklist Civilian Version	[30,32]
Pittsburgh Sleep Quality Scale (PSQI-A[c] for PTSD)	[30]
Other	
12-item Short Form Survey	[43]
Flinders Fatigue Scale	[43]
Clinical Global Impression–Improvement Scale	[30]
West Haven-Yale Multidimensional Pain Inventory	[32]
Pain Disability Index	[32]
Marijuana smoking history	[34]
Timeline follow-back interview	[34]

[a]PHQ: Patient Health Questionnaire.

[b]PTSD: posttraumatic stress disorder.

[c]PSQI-A: Pittsburgh Sleep Quality Index–Addendum.

Discussion

Principal Findings

This study provides the first comprehensive review of published studies and a framework for the design engineering and clinical implementation and evaluation of mobile apps for sleep disturbance. Despite the availability of over 500 [50] sleep mobile apps in commercial app stores (Apple App and Google Play Store), our review identified 15 papers assessing the design, implementation, and evaluation of 8 apps, only one of which was available to download on commercial app stores [32,33,40,45-47]. This means that less than 1% of all commercially available sleep apps have any published data on these aspects. Of the 15 papers, implementation metrics were reported in 10 papers and treatment outcomes were evaluated in 11 papers. Despite the potential of mHealth, there has been a small number of studies, a lack of standardization in design engineering approaches and clinical implementation assessments, and few comprehensive clinical evaluations.

Design Engineering

For the increased utilization and adoption of mHealth apps, these technologies must be designed for people who will use them—both end users and multidisciplinary stakeholders [51-54]. Our review indicates that although some apps utilized best practice design approaches, for example, user-centered and multidisciplinary methods, approximately half of the apps did not report their design approach. We have previously shown how people with insomnia have unique user needs and preferences for sleep mobile apps that can drive engagement [42]; however, only 2 of the 8 apps reported any end-user involvement. Multidisciplinary teams are particularly crucial in a domain such as sleep, where various stakeholders (eg, clinical psychologists, sleep clinicians, psychiatrists) tend to be involved in patient care; however, only 2 apps reported this approach. Although these best practice design approaches may be an unspoken rule in app development, transparency in the reporting of these approaches is important in encouraging clinicians to be able to recommend such apps designed to sustain engagement.

Clinical Implementation

There were few comprehensive evaluations of implementation, with only 3 out of 8 apps reporting more than 2 implementation metrics. Poor implementation can limit adoption or engagement, particularly in an uncontrolled and real-world setting, which in turn limits effectiveness. Our results suggest that although most apps had papers reporting some implementation outcomes, very few conducted a comprehensive exploration. For instance, although there is support for Sleepcare's efficacy, there was a lack of a substantial assessment for its implementation. The National Health Service in the United Kingdom highlights the importance of implementation in the acknowledgment that their previous failure of digital health technology deployment was attributed to rushed and inadequate implementation [55]. Moreover, reporting of implementation processes may enable replication and reduce the gap between research and practice [56].

There was great heterogeneity among implementation studies in the conceptualization of implementation metrics. Given the inconsistency in terminology, we developed our own definitions to facilitate cross-study comparisons. This language incongruence highlights the need for a taxonomy of implementation metrics to clearly delineate key variables [57,58]. Similarly, there was a lack of standardized implementation metrics ranging from qualitative methods to nonstandardized quantitative surveys. This further contributes to blurring among constructs. When standardized measures were used, they were not specifically designed for mobile or sleep disturbances. For instance, the System Usability Scale was one of the most frequently used implementation scales in the identified studies [59]. Although it is a well-researched measure with good psychometrics [59], it is not designed for mobile or sleep use. A taxonomy of implementation outcomes with standardized tools for sleep mobile apps can advance the measurement and understanding of implementation processes.

Despite the importance of privacy and regulatory oversight in digital technology, only 3 papers mentioned privacy, and no papers considered mHealth regulations. Mobile devices collect a large amount of behavioral and health care data, raising concerns regarding mHealth app quality and safety [60-62]. Most mHealth apps do not have a transparent privacy policy, leaving end users unaware of what data are collected, how data are transferred, where data are stored, and with whom the data are shared [63]. Given that data security is a key concern among health care providers when recommending mHealth apps [64], transparent privacy policies and further regulatory oversight from bodies such as the FDA can combat this issue.

Clinical Evaluation

Of the evaluation studies, we identified 2 RCTs, only one of which was adequately powered according to the primary paper [44]. Most of the apps (6/8, 75%) delivered full or modified CBT-I (SRT or BTi), a well-established intervention where an evidence base already exists for face-to-face and internet-enabled delivery. CBT-I has been recommended as a first-line therapy for insomnia, given its substantial clinical base [8-10], and has shown comparable efficacy when delivered via the internet [65]. Although a previous systematic review demonstrated support for the efficacy of mobile phone interventions for sleep disorders and sleep quality, this review only included studies of 4 mobile apps [66]. In this review, only 2 of these 4 apps were included. One was excluded as it was a multimodal intervention and the other was targeted toward older adults. There is an evident need for methodologically robust and adequately powered studies assessing the effectiveness of mHealth apps for sleep disturbance. Nevertheless, mobile-delivered CBT-I has potential, given the therapy's existing evidence base across various modalities.

Most apps described in this review were prototype versions, with only one app being matured to released (CBT-I Coach).

This is aligned with a greater number of identified studies primarily focusing on earlier stages of development, for example, design or implementation, with some preliminary evaluations of efficacy. Of the apps identified in this review, only 1 app (Sleepcare) has the support of efficacy from an adequately powered RCT [44]. Despite this, there are no implementation studies for Sleepcare, and although acceptance is included as a measure in the paper, the results are not reported. Conversely, several papers described the design and implementation of the CBT-I Coach, but no full-scale efficacy evaluation was identified. However, CBT-I Coach is the only app available in commercial app stores.

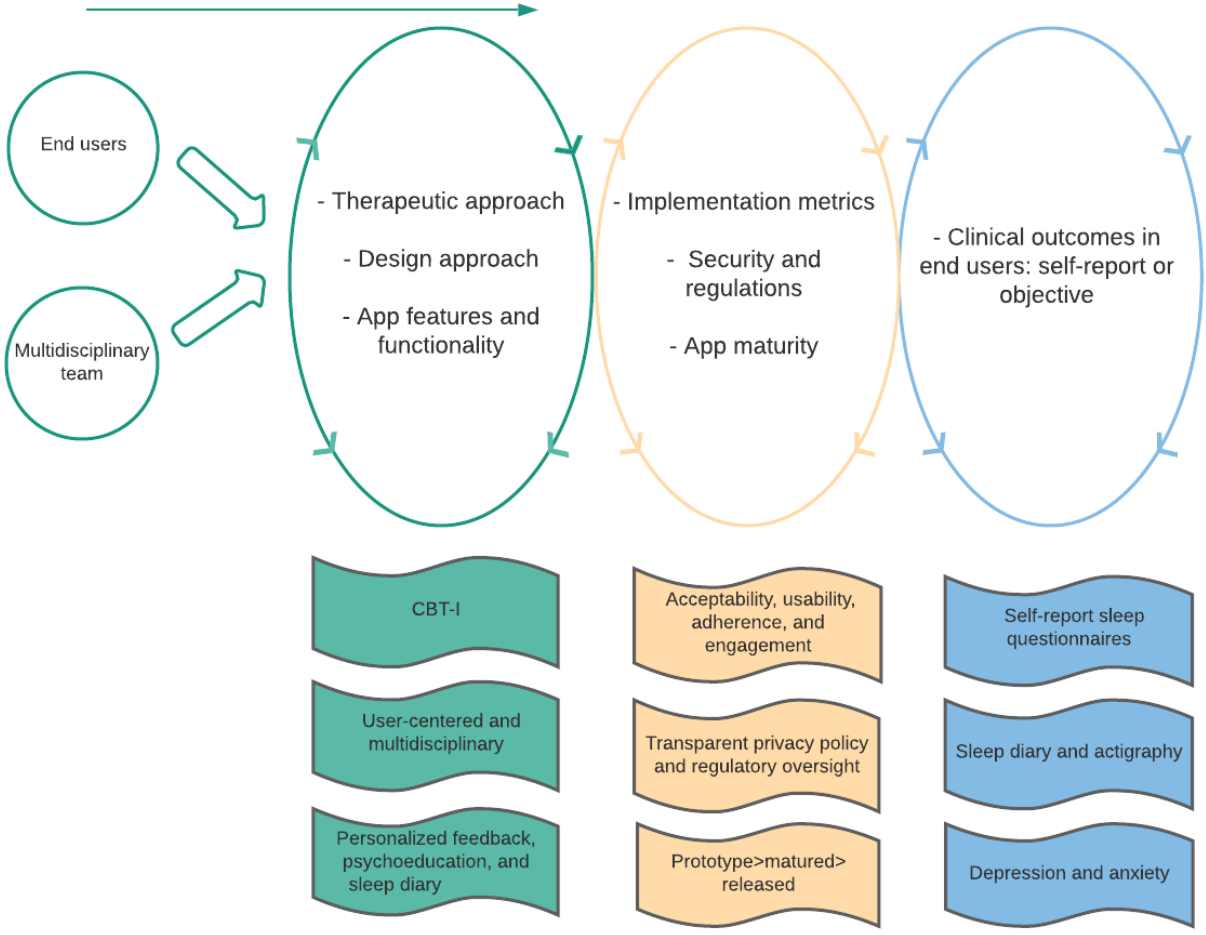
Although it might be thought that a more mature app would have more cumulative evidence for its design, implementation, and evaluation, this does not prove to be the case. This mismatch in the maturity of sleep apps and the levels of available evidence ultimately reflects the unregulated nature of mHealth app development and deployment. To exacerbate the problem, the commercial app marketplace allows developers and researchers to freely release apps into these stores, which serve as the main app repository for consumers. Evidently, there is a need for a standardized set of evidence-based criteria for researchers to meet before making commercially available apps.

Ultimately, the lack of standardization in the evidence and reporting for the design engineering and clinical implementation and evaluation of mHealth apps for sleep disturbance stresses the need for a comprehensive framework to guide researchers and app developers. A recent systematic review highlighted the wide heterogeneity among the different published criteria for the assessment of mHealth apps, with 38 main classes of criteria [19]. Guidelines specific to the development and assessment of apps for sleep disturbance are particularly scarce. Establishing an extensive and standardized framework for mobile apps for sleep disturbance may lead to improved existing tools and the development of successful, high-quality, and effective apps.

mHealth App Framework for Sleep Disturbance

As a first step, we developed a high-level framework based on the findings of this study to guide the design engineering and clinical implementation and evaluation of apps for sleep disturbance. The findings are summarized in Figure 2.

Figure 2. Framework for the design engineering and clinical implementation and evaluation of apps for sleep disturbance. CBT-I: cognitive behavioral therapy for insomnia.



Although several frameworks exist for conditions such as PTSD [22], bipolar disorder [23], and hypertension [24], there are no frameworks for the development of sleep apps. Each of these frameworks similarly address design engineering, clinical implementation, and evaluation. For instance, the framework for bipolar disorder also notes the importance of designing with

end users and multidisciplinary teams, addressing security and regulations with standards consistent with HIPAA, and evaluation with end users [23]. These frameworks were developed through a combination of lessons learned from firsthand app development, best practice principles, and theory-based design models. The framework in this study triangulates the findings from this review of digital sleep interventions to these previous frameworks augmented by our firsthand experience with the development of an app for insomnia [43,67]. Future work may consider adapting theory-based design models for sleep disturbance and integrating them into this framework.

Several reviews of commercial sleep apps have demonstrated a lack of validated sleep measurement algorithms [68], evidence-based principles for insomnia management [69], behavior change constructs [70], and overall low quality of functionality and content based on established app assessment criteria [71,72]. Evidently, commercial development of apps has severely outpaced academic research, putting their trustworthiness in question [73]. Our systematic search identified 13 clinical trial registrations, of which 6 were mobile apps not included in our systematic review as there were no available publications. Although partly attributable to the infancy of the mHealth field, there is still a necessity for timely and increased efforts of mobile sleep apps to progress to clinical evaluations. Collaboration between academia and industry may offer an opportunity to work together in developing scientifically rigorous solutions while keeping pace with the rapidly evolving app market.

This review has several limitations. First, we included English language publications only, which introduces publication bias, particularly given that these papers tended to originate from high-income countries such as the United States and Australia. Second, given that data extraction was based on the included studies only and that the mobile apps were not downloaded by the authors, some information such as app features and design approaches was not always clear or available. Third, given that our study focused on apps for sleep disturbance and did not include mHealth apps with multimodal interventions, including sleep, the inferences from this study may not extend to all sleep apps.

Conclusions

This is the first review to evaluate the design engineering and clinical implementation and evaluation of apps designed for sleep disturbance. It was found that despite a plethora of sleep apps available, there is limited research and a lack of standardization in the evidence base for the design, implementation, and evaluation of apps for sleep disturbance. Few apps had evidence for the use of best practice design approaches. Implementation assessments lacked standardization and consistency in implementation metrics used, and very few comprehensive efficacy evaluations were identified.

For the future development of engaging and evidence-based apps for sleep disturbance, we have developed a framework to guide the development and deployment process. The framework aims to address the need for (1) increased application and reporting of best practice design approaches, for example, user-centered and multidisciplinary teams; (2) comprehensive implementation assessments involving multiple metrics, tools validated for sleep, and privacy and regulatory considerations; and (3) rigorous evaluations of clinical efficacy. Collaboration between academia and the industry may facilitate the development of evidence-based apps in the fast-paced mHealth technology environment.

Conflicts of Interest

MA, CG, RC, RG, and NG are named on 2 provisional patents for the SleepFix app.

Multimedia Appendix 1
Example search strategy.
[DOCX File , 12 KB - jmir_v23i2e24607_app1.docx]

References

1. Adams RJ, Appleton SL, Taylor AW, Gill TK, Lang C, McEvoy RD, et al. Sleep health of Australian adults in 2016: results of the 2016 Sleep Health Foundation national survey. Sleep Health 2017 Feb;3(1):35-42. [doi: 10.1016/j.sleh.2016.11.005] [Medline: 28346149]
2. Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds CF. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. World Psychiatry 2013 Jun;12(2):137-148 [FREE Full text] [doi: 10.1002/wps.20038] [Medline: 23737423]
3. Baglioni C, Battagliese G, Feige B, Spiegelhalder K, Nissen C, Voderholzer U, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J Affect Disord 2011 Dec;135(1-3):10-19. [doi: 10.1016/j.jad.2011.01.011] [Medline: 21300408]
4. Taylor DJ, Lichstein KL, Durrence HH, Reidel BW, Bush AJ. Epidemiology of insomnia, depression, and anxiety. Sleep 2005 Nov;28(11):1457-1464. [doi: 10.1093/sleep/28.11.1457] [Medline: 16335332]
5. Khan MS, Aouad R. The effects of insomnia and sleep loss on cardiovascular disease. Sleep Med Clin 2017 Jun;12(2):167-177. [doi: 10.1016/j.jsmc.2017.01.005] [Medline: 28477772]

6. Vgontzas AN, Liao D, Bixler EO, Chrousos GP, Vela-Bueno A. Insomnia with objective short sleep duration is associated with a high risk for hypertension. Sleep 2009 Apr;32(4):491-497 [FREE Full text] [doi: 10.1093/sleep/32.4.491] [Medline: 19413143]
7. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2010 Feb;33(2):414-420 [FREE Full text] [doi: 10.2337/dc09-1124] [Medline: 19910503]
8. van Straten A, van der Zweerde T, Kleiboer A, Cuijpers P, Morin CM, Lancee J. Cognitive and behavioral therapies in the treatment of insomnia: a meta-analysis. Sleep Med Rev 2018 Apr;38:3-16. [doi: 10.1016/j.smrv.2017.02.001] [Medline: 28392168]
9. Jindal RD. Cognitive behavioral therapy alone and with medication for persistent insomnia. J Am Med Assoc 2009 Sep 09;302(10):1053. [doi: 10.1001/jama.2009.1282] [Medline: 19738086]
10. Mitchell MD, Gehrman P, Perlis M, Umscheid CA. Comparative effectiveness of cognitive behavioral therapy for insomnia: a systematic review. BMC Fam Pract 2012 May 25;13:40 [FREE Full text] [doi: 10.1186/1471-2296-13-40] [Medline: 22631616]
11. 325,000 mobile health apps available in 2017. Research2Guidance. 2017. URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017 [accessed 2020-07-07]
12. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. JMIR Mhealth Uhealth 2017 Mar 14;5(3):35 [FREE Full text] [doi: 10.2196/mhealth.6522] [Medline: 28292740]
13. Mateo GF, Font EG, Grau CF, Carreras XM. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
14. Creber RMM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
15. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
16. Firth J, Torous J. Smartphone apps for schizophrenia: a systematic review. JMIR Mhealth Uhealth 2015 Nov 06;3(4):102 [FREE Full text] [doi: 10.2196/mhealth.4930] [Medline: 26546039]
17. Grundy Q, Chiu K, Held F, Continella A, Bero L, Holz R. Data sharing practices of medicines related apps and the mobile ecosystem: traffic, content, and network analysis. BMJ 2019 Mar 20;364:920 [FREE Full text] [doi: 10.1136/bmj.l920] [Medline: 30894349]
18. Armontrout J, Torous J, Fisher M, Drogin E, Gutheil T. Mobile mental health: navigating new rules and regulations for digital tools. Curr Psychiatry Rep 2016 Oct;18(10):91. [doi: 10.1007/s11920-016-0726-x] [Medline: 27553979]
19. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 01;25(8):1089-1098 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
20. Buijink AWG, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med 2013 Jun;18(3):90-92. [doi: 10.1136/eb-2012-100885] [Medline: 22923708]
21. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. Telemed J E Health 2015 Dec;21(12):1038-1041. [doi: 10.1089/tmj.2015.0002] [Medline: 26171663]
22. Schellong J, Lorenz P, Weidner K. Proposing a standardized, step-by-step model for creating post-traumatic stress disorder (PTSD) related mobile mental health apps in a framework based on technical and medical norms. Eur J Psychotraumatol 2019;10(1):1611090 [FREE Full text] [doi: 10.1080/20008198.2019.1611090] [Medline: 31143412]
23. Bauer AM, Hodsdon S, Bechtel JM, Fortney JC. Applying the principles for digital development: case study of a smartphone app to support collaborative care for rural patients with posttraumatic stress disorder or bipolar disorder. J Med Internet Res 2018 Jun 06;20(6):10048 [FREE Full text] [doi: 10.2196/10048] [Medline: 29875085]
24. Duan H, Wang Z, Ji Y, Ma L, Liu F, Chi M, et al. Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR Mhealth Uhealth 2020 Feb 25;8(2):14466 [FREE Full text] [doi: 10.2196/14466] [Medline: 32130161]
25. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
26. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
27. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

28. WHO. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. WHO 2016:1-144 [FREE Full text] [doi: 10.1017/cbo9780511712074.003]
29. De Geest S, Sabaté E. Adherence to long-term therapies: evidence for action. Eur J Cardiovascular Nursing 2016 Jun 22;2(4):323-323. [doi: 10.1016/s1474-5151(03)00091-4]
30. Pulantara IW, Parmanto B, Germain A. Clinical feasibility of a just-in-time adaptive intervention app (iREST) as a behavioral sleep treatment in a military population: feasibility comparative effectiveness study. J Med Internet Res 2018 Dec 07;20(12):10124 [FREE Full text] [doi: 10.2196/10124] [Medline: 30530452]
31. Pulantara IW, Parmanto B, Germain A. Development of a just-in-time adaptive mhealth intervention for insomnia: usability study. JMIR Hum Factors 2018 May 17;5(2):21 [FREE Full text] [doi: 10.2196/humanfactors.8905] [Medline: 29773529]
32. Reilly ED, Robinson SA, Petrakis BA, Kuhn E, Pigeon WR, Wiener RS, et al. Mobile app use for insomnia self-management: pilot findings on sleep outcomes in veterans. Interact J Med Res 2019 Jul 24;8(3):12408 [FREE Full text] [doi: 10.2196/12408] [Medline: 31342904]
33. Koffel E, Kuhn E, Petsoulis N, Erbes CR, Anders S, Hoffman JE, et al. A randomized controlled pilot study of CBT-I Coach: feasibility, acceptability, and potential impact of a mobile phone application for patients in cognitive behavioral therapy for insomnia. Health Informatics J 2018 Mar;24(1):3-13 [FREE Full text] [doi: 10.1177/1460458216656472] [Medline: 27354394]
34. Babson KA, Ramo DE, Baldini L, Vandrey R, Bonn-Miller MO. Mobile app-delivered cognitive behavioral therapy for insomnia: feasibility and initial efficacy among veterans with cannabis use disorders. JMIR Res Protoc 2015 Jul 17;4(3):87 [FREE Full text] [doi: 10.2196/resprot.3852] [Medline: 26187404]
35. Shirazi A, Clawson J, Hassanpour Y. Int J Human-Computer Stud. 2013. URL: https://doi.org/10.1016/j.ijhcs.2013.03.001 [accessed 2021-01-22]
36. Bhat S, Ferraris A, Gupta D, Mozafarian M, DeBari VA, Gushway-Henry N, et al. Is there a clinical role for smartphone sleep apps? Comparison of sleep cycle detection by a smartphone application to polysomnography. J Clin Sleep Med 2015 Jul 15;11(7):709-715 [FREE Full text] [doi: 10.5664/jcsm.4840] [Medline: 25766719]
37. Fino E, Plazzi G, Filardi M, Marzocchi M, Pizza F, Vandi S, et al. (Not so) Smart sleep tracking through the phone: findings from a polysomnography study testing the reliability of four sleep applications. J Sleep Res 2020 Feb;29(1):12935. [doi: 10.1111/jsr.12935] [Medline: 31674096]
38. Scott H, Lack L, Lovato N. A pilot study of a novel smartphone application for the estimation of sleep onset. J Sleep Res 2018 Feb;27(1):90-97 [FREE Full text] [doi: 10.1111/jsr.12575] [Medline: 28656615]
39. Baron KG, Duffecy J, Reid K, Begale M, Caccamo L. Technology-assisted behavioral intervention to extend sleep duration: development and design of the sleep bunny mobile app. JMIR Ment Health 2018 Jan 10;5(1):3 [FREE Full text] [doi: 10.2196/mental.8634] [Medline: 29321122]
40. Bauer J, Consolvo S, Greenstein B. ShutEy: encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery; 2012 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; ; Austin, Texas, USA; May, 2012; Austin Texas USA p. 1401-1410. [doi: 10.1145/2207676.2208600]
41. Kang S, Kang JM, Cho S, Ko K, Lee YJ, Lee H, et al. Cognitive behavioral therapy using a mobile application synchronizable with wearable devices for insomnia treatment: a pilot study. J Clin Sleep Med 2017 Apr 15;13(4):633-640 [FREE Full text] [doi: 10.5664/jcsm.6564] [Medline: 28162145]
42. Aji M, Gordon C, Peters D, Bartlett D, Calvo RA, Naqshbandi K, et al. Exploring user needs and preferences for mobile apps for sleep disturbance: mixed methods study. JMIR Ment Health 2019 May 24;6(5):13895 [FREE Full text] [doi: 10.2196/13895] [Medline: 31127714]
43. Aji M, Glozier N, Bartlett D, Peters D, Calvo RA, Zheng Y, et al. A feasibility study of a mobile app to treat insomnia. Transl Behav Med 2020 Mar 30:-. [doi: 10.1093/tbm/ibaa019] [Medline: 32227087]
44. Horsch CH, Lancee J, Griffioen-Both F, Spruit S, Fitrianie S, Neerincx MA, et al. Mobile phone-delivered cognitive behavioral therapy for insomnia: a randomized waitlist controlled trial. J Med Internet Res 2017 Apr 11;19(4):70 [FREE Full text] [doi: 10.2196/jmir.6524] [Medline: 28400355]
45. Kuhn E, Weiss BJ, Taylor KL, Hoffman JE, Ramsey KM, Manber R, et al. CBT-I coach: a description and clinician perceptions of a mobile app for cognitive behavioral therapy for insomnia. J Clin Sleep Med 2016 Apr 15;12(4):597-606 [FREE Full text] [doi: 10.5664/jcsm.5700] [Medline: 26888586]
46. Miller KE, Kuhn E, Owen JE, Taylor K, Yu JS, Weiss BJ, et al. Clinician perceptions related to the use of the CBT-I coach mobile app. Behav Sleep Med 2019;17(4):481-491. [doi: 10.1080/15402002.2017.1403326] [Medline: 29120247]
47. Omeogu C, Shofer F, Gehrman P, Green-McKenzie J. Efficacy of a mobile behavioral intervention for workers with insomnia. J Occup Environ Med 2020 Mar;62(3):246-250. [doi: 10.1097/JOM.0000000000001819] [Medline: 31977925]
48. Inal Y, Wake JD, Guribye F, Nordgreen T. Usability evaluations of mobile mental health technologies: systematic review. J Med Internet Res 2020 Jan 06;22(1):15337 [FREE Full text] [doi: 10.2196/15337] [Medline: 31904579]
49. Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 01;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]

50. Ko PT, Kientz JA, Choe EK, Kay M, Landis CA, Watson NF. Consumer sleep technologies: a review of the landscape. J Clin Sleep Med 2015 Dec 15;11(12):1455-1461 [FREE Full text] [doi: 10.5664/jcsm.5288] [Medline: 26156958]
51. Maguire M. Methods to support human-centred design. Int J Human-Computer Stud. URL: https://doi.org/10.1006/ijhc.2001.0503 [accessed 2021-01-22]
52. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
53. Shah SGS, Robinson I, AlShawi S. Developing medical device technologies from users' perspectives: a theoretical framework for involving users in the development process. Int J Technol Assess Health Care 2009 Oct;25(4):514-521. [doi: 10.1017/S0266462309990328] [Medline: 19845981]
54. Ben-Zeev D, Schueller SM, Begale M, Duffecy J, Kane JM, Mohr DC. Strategies for mHealth research: lessons from 3 mobile intervention studies. Adm Policy Ment Health 2015 Mar;42(2):157-167 [FREE Full text] [doi: 10.1007/s10488-014-0556-2] [Medline: 24824311]
55. NHS. Five year forward view. National Health Service. United Kingdom: NHS; 2014. URL: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf
56. Pfadenhauer LM, Mozygemba K, Gerhardus A, Hofmann B, Booth A, Lysdahl KB, et al. Context and implementation: a concept analysis towards conceptual maturity. Z Evid Fortbild Qual Gesundhwes 2015;109(2):103-114. [doi: 10.1016/j.zefq.2015.01.004] [Medline: 26028447]
57. Dubad M, Winsper C, Meyer C, Livanou M, Marwaha S. A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people. Psychol Med 2018 Jan;48(2):208-228. [doi: 10.1017/S0033291717001659] [Medline: 28641609]
58. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
59. Brooke J. SUS: a 'Quick and Dirty' usability scale. In: Usability Evaluation In Industry. Florida: CRC Press; 1996.
60. Powell AC, Landman AB, Bates DW. In search of a few good apps. J Am Med Assoc 2014 May 14;311(18):1851-1852. [doi: 10.1001/jama.2014.2564] [Medline: 24664278]
61. Kargl F, van der Heijden RW, Erb B, Bösch C. Privacy in mobile sensing. Digital Phenotyping and Mobile Sensing 2019:3-12. [doi: 10.1007/978-3-030-31620-4_1]
62. Parker L, Bero L, Gillies D, Raven M, Grundy Q. The "Hot Potato" of mental health app regulation: a critical case study of the australian policy arena. Int J Health Policy Manag 2019 Mar 01;8(3):168-176 [FREE Full text] [doi: 10.15171/ijhpm.2018.117] [Medline: 30980633]
63. O'Loughlin K, Neary M, Adkins EC, Schueller SM. Reviewing the data security and privacy policies of mobile apps for depression. Internet Interv 2019 Mar;15:110-115 [FREE Full text] [doi: 10.1016/j.invent.2018.12.001] [Medline: 30792962]
64. Aguilera A, Muench F. There's an app for that: information technology applications for cognitive behavioral practitioners. Behav Ther (N Y N Y) 2012 Apr;35(4):65-73 [FREE Full text] [Medline: 25530659]
65. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016 Dec;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
66. Shin JC, Kim J, Grigsby-Toussaint D. Mobile phone interventions for sleep disorders and sleep quality: systematic review. JMIR Mhealth Uhealth 2017 Sep 07;5(9):131 [FREE Full text] [doi: 10.2196/mhealth.7244] [Medline: 28882808]
67. Aji M, Gordon C, Peters D, Bartlett D, Calvo RA, Naqshbandi K, et al. Exploring user needs and preferences for mobile apps for sleep disturbance: mixed methods study. JMIR Ment Health 2019 May 24;6(5):13895 [FREE Full text] [doi: 10.2196/13895] [Medline: 31127714]
68. Ong AA, Gillespie MB. Overview of smartphone applications for sleep analysis. World J Otorhinolaryngol Head Neck Surg 2016 Mar;2(1):45-49 [FREE Full text] [doi: 10.1016/j.wjorl.2016.02.001] [Medline: 29204548]
69. Yu JS, Kuhn E, Miller KE, Taylor K. Smartphone apps for insomnia: examining existing apps' usability and adherence to evidence-based principles for insomnia management. Transl Behav Med 2019 Jan 01;9(1):110-119. [doi: 10.1093/tbm/iby014] [Medline: 30590862]
70. Grigsby-Toussaint DS, Shin JC, Reeves DM, Beattie A, Auguste E, Jean-Louis G. Sleep apps and behavioral constructs: a content analysis. Prev Med Rep 2017 Jun;6:126-129 [FREE Full text] [doi: 10.1016/j.pmedr.2017.02.018] [Medline: 28316907]
71. Choi YK, Demiris G, Lin S, Iribarren SJ, Landis CA, Thompson HJ, et al. Smartphone applications to support sleep self-management: review and evaluation. J Clin Sleep Med 2018 Oct 15;14(10):1783-1790 [FREE Full text] [doi: 10.5664/jcsm.7396] [Medline: 30353814]
72. Leigh S, Ouyang J, Mimnagh C. Effective? Engaging? Secure? Applying the ORCHA-24 framework to evaluate apps for chronic insomnia disorder. Evid Based Ment Health 2017 Nov;20(4):20. [doi: 10.1136/eb-2017-102751] [Medline: 28947676]
73. Wykes T, Schueller S. Why reviewing apps is not enough: transparency for trust (T4T) principles of responsible health app marketplaces. J Med Internet Res 2019 May 02;21(5):12390 [FREE Full text] [doi: 10.2196/12390] [Medline: 31045497]

Abbreviations

BTi: behavioral therapy for insomnia
CBT-I: cognitive behavioral therapy for insomnia
FDA: Food and Drug Administration
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health
PTSD: posttraumatic stress disorder
SRT: sleep restriction therapy

Edited by G Eysenbach; submitted 27.09.20; peer-reviewed by R Krukowski, Z Aghaei; comments to author 09.10.20; revised version received 12.10.20; accepted 15.01.21; published 17.02.21.

Please cite as:
Aji M, Gordon C, Stratton E, Calvo RA, Bartlett D, Grunstein R, Glozier N
Framework for the Design Engineering and Clinical Implementation and Evaluation of mHealth Apps for Sleep Disturbance: Systematic Review
J Med Internet Res 2021;23(2):e24607
URL: http://www.jmir.org/2021/2/e24607/
doi:10.2196/24607
PMID:33595441

Original Paper

How Social Media Use at Work Affects Improvement of Older People's Willingness to Delay Retirement During Transfer From Demographic Bonus to Health Bonus: Causal Relationship Empirical Study

Yiming Ma[1], PhD; Changyong Liang[1], PhD; Dongxiao Gu[1,2], PhD; Shuping Zhao[1], PhD; Xuejie Yang[1], PhD; Xiaoyu Wang[3], PhD

[1]School of Management, Hefei University of Technology, Hefei, China

[2]Key Laboratory of Process Optimization and Intelligent Decision-making of Ministry of Education, Hefei, China

[3]The First Affiliated Hospital, Anhui University of Traditional Chinese Medicine, Hefei, China

Corresponding Author:
Dongxiao Gu, PhD
School of Management
Hefei University of Technology
193 Tunxi Road
Hefei, 230009
China
Phone: 86 0551 62904956
Email: gudongxiao@hfut.edu.cn

Abstract

Background: With the increased older population in China and the subsequent reduced labor force, the "demographic bonus" is disappearing. The Chinese government proposed a Healthy China strategy in 2017. The transfer of the demographic bonus to a "health bonus" extended the working life of people and reduced the negative impact of the population's aging on the labor force structure.

Objective: This research focuses on the effect of older workers' social media usage at work on their work ability (related to both physical and mental health) and thus their willingness to delay retirement.

Methods: The questionnaire respondents were older than 55 years, and they obtained the questionnaire from social media, from June to July 2018. A total of 1020 valid questionnaires were collected, and SmartPLS 3.28 (SmartPLS GmbH) was used to analyze the data. Effects were analyzed using 2-tailed *t* tests.

Results: (1) Use of social media at work can improve information support (t_{14}=13.318, P<.001), emotional support (t_{14}=13.184, P<.001), and self-efficacy (t_{14}=6.364, P<.001) for older people; (2) information support is the main factor affecting the self-efficacy of older workers (t_{14}=23.304, P<.001), as compared with emotional support (t_{14}=1.799, P=0.07); (3) the impacts of emotional support on work ability (t_{14}=8.876, P<.001) and work stress (t_{14}=9.545, P<.001) are generally higher than those of information support (t_{14}=4.394, P<.001; t_{14}=5.002, P<.001); (4) self-efficacy has an impact on work ability (t_{14}=5.658, P<.001) and work stress (t_{14}=4.717, P<.001); and (5) the impacts of work ability (t_{14}=8.586, P<.001) and work stress (t_{14}=8.579, P<.001) on retirement willingness are greater than those of emotional support (t_{14}=2.112, P=.04) and information support (t_{14}=4.314, P<.001).

Conclusions: Our study confirms that the use of social media at work has a positive impact on older workers. Based on the findings, we have put forward proposals to extend people's working lives and help governments implement health bonus policies. In the future, we will compare the different values of willingness to delay retirement among older people in different occupations and different cultures.

(J Med Internet Res 2021;23(2):e18264) doi:10.2196/18264

KEYWORDS

social media; older workers; social support; work ability; delayed retirement

Introduction

Background

The problems associated with an aging population have become a worldwide challenge, affecting multiple countries and regions. According to a World Health Organization report from late 2017, the proportion of the world's population older than 60 years will double by 2050, increasing from 11% to 22%. According to estimates, by 2050, the absolute number of people older than 60 years will increase from 605 million to 2 billion [1]. This aging population means that the "demographic bonus" has disappeared in a large number of countries. The demographic bonus refers to the situation where a falling birth rate reduces the number of minors, reduces the burden on the family, and forms a relatively rich labor force, which is conducive to economic development [2]. However, the aging population also means that the labor force's share of the population is falling, the dependency ratio is rising (ie, the dependency of older people and children), and the economy is developing slowly because of the shortage of labor. Therefore, it is necessary to find new sources of labor to reduce the negative impact of a labor shortage.

While high-income countries have the challenges of an aging population and declining demographic bonus, they also have a population of older adults (older than 60 years) who are healthier [3], with better working capabilities, compared with those in low-income countries. Therefore, developing human resources among older adults and promoting the generation of a second round of a demographic bonus has become a problem that needs to be solved by all countries troubled by the challenges of an aging population [4]. The Chinese government proposed a Healthy China strategy in 2017, promoting the transformation of the demographic bonus to a "health bonus," and improving working life, which will reduce the aging population's negative impact on the structure of the labor force. The health bonus refers to the situation where labor productivity is improved to alleviate the labor shortage caused by the aging population; this provides a foundation for the healthy growth of the population of older adults who have delayed retirement, as well as adults of childbearing age, by improving the health of the workforce [5].

Determining how to maintain and improve the physical and mental health level of older workers and optimize the working ability and work experience of older adults has become an important prerequisite to realize the health bonus. Compared with other age groups, older individuals are more susceptible to psychological and mental problems [6]. Effective social support can prevent and alleviate psychological problems and improve the physiological health of older adults, to a certain extent [7]. Therefore, increasing older adults' access to social support is of great significance to extending their working life and delaying their retirement. Especially for older working adults, different types of social support can not only help older workers to improve their physical and mental health but also help them to acquire work-related knowledge and skills and prevent the decline in work that is often caused by aging.

With the development of information and communications technology and the use of social media, access to social support is no longer realized through a single offline channel but can be realized through interactive online and offline channels [7]. Thus, using social media can help older people obtain social support. At present, social media has become an integral part of work. Some people think that the use of social media at work can have negative effects (such as reducing productivity and increasing disturbances) [8]. However, recent studies have shown that social media plays a positive role in work. By using social media at work, people can quickly assign work tasks and report work status and consultation, and workers can obtain different types of social support, which can have a positive impact on people in many ways. This social support can establish and strengthen connections among colleagues, help workers collect professional information, and promote knowledge and resource sharing [9]. Moreover, these behaviors can also promote people's achievement of self-efficacy and improve their self-confidence at work [10]. In addition, non–work-related social media use behaviors (such as entertainment behaviors on social media) can reduce work stress and psychological problems [9]. And interactions on social media related to health information (such as competitive step counting) can also promote exercise around the workplace and maintain people's level of health [11]. In conclusion, the existing research has suggested that the use of social media at work can improve workers' mental health and work ability.

However, the existing research on social media mainly targets "ordinary" employees, resulting in a gap in research on older workers (older than 60 years). Research targeting older workers focuses on their working status, with a gap in research on their willingness to delay retirement. For these older workers, the main factors affecting their retirement are not only economic factors but also health level and job satisfaction [12,13]. The use of social media at work can improve the physical and mental health of older workers to a certain extent and improve their working ability. Furthermore, it can improve the performance of older workers and make them more likely to achieve job satisfaction. Thus, the use of social media at work can affect older workers' willingness to delay retirement. Currently, there is a lack of empirical evidence to confirm that the use of social media at work can affect delayed retirement. In addition, further understanding the demands of employees older than 60 years, as well as how to enhance their ability to work and increase their willingness to delay retirement, promotes the development of this human resource and is of great economic and social value and important theoretical significance. Therefore, this paper puts forward the following questions to explore the impact of social media use at work on the intention of older workers to delay retirement:

1. How does social media use at work improve older workers' physical and mental health, and how does it affect work ability and job burnout?
2. Will social media use at work affect the expected working duration of older workers in the future?

Based on theories relating social support and work ability, this paper studies the work ability and working duration of older adults from the perspective of social media. This paper narrows

the focus from all sectors of the community to the sector of older workers, so as to actively improve their workplace conditions and create a better environment for them, thus promoting the health bonus.

Literature Review and Hypotheses

Social Media at Work

Social media is a highly interactive platform based on information and communication technology. Individuals and communities can share, create, discuss, and modify user-generated content through these platforms such as WeChat, Facebook, Twitter, and QQ [14]. The functions of social media include identity, conversations, sharing, presence, relationships, reputation, and groups [14]. The use of social media at work has become a common phenomenon [15]. Some scholars believe that the use of social media at work will distract employees' attention and lead to a decline in work efficiency [8]. However, from a social support perspective, social media used at work is beneficial [16].

Social support is defined as the assistance an individual can access from the social resources of his or her social network [17]. Social support can be divided into four categories: material, emotional, information, and companion support [17]. The social support obtained at work mainly includes information support and emotional support [18]. These different types of social support are important for people to access work-related resources and improve their work-related abilities. Through the use of social media at work, people can more easily access various types of social support. Therefore, we believe that the positive effects of social media use at work include the following aspects:

First, the use of social media at work can increase social support for employees [18]. This kind of social support includes emotional and information support. Social media can make it easier for employees to connect and interact with colleagues [18]. This kind of interaction can facilitate the sharing of work experiences and exchanges of knowledge among employees (information support) [19]. For older people, the use of social media can help them maintain their existing relationships, while the process of learning the method to use social media is also a process of acquiring social support (information and emotional support) [20]. In addition, older people tend to need more attention and support [21]; for older workers, this kind of attention and support is extraordinarily important (emotional support). The characteristics of social media, such as conversations, sharing, presence, and relationships, can increase the acquisition of emotional support [20]. Based on the literature cited above, we believe that the use of social media at work can improve the level of social support for older employees, including emotional support and information support. Therefore, we propose the following hypotheses:

> H1: The use of social media at work can improve emotional support for older workers.
>
> H2: The use of social media at work can improve information support for older workers.

Second, social support gained through the use of social media can also increase self-efficacy. Self-efficacy refers to an individual's belief in his or her capacity to execute behaviors necessary to achieve specific performance outcomes [22]. At work, the level of self-efficacy often determines the individual's work ability and work performance. Factors that affect self-efficacy include knowledge, skills, experience, and social support [22]. Social support can significantly regulate an individual's self-efficacy [23]. On the one hand, social support can promote the acquisition of new knowledge and skills (information support), thereby improving self-efficacy [23]. On the other hand, emotional support can also affect people's self-efficacy. Self-efficacy is constantly changing on a daily basis [24]. The change in self-efficacy is often due to changes in interpersonal relationships [25]. As an important tool for maintaining an individual's interpersonal relationships, social media can help people maintain their social networks in a working environment and help them give or get emotional support. In addition, for older people, the use of social media directly affects an individual's sense of self-efficacy. Involvement in and use of social media can heighten an individual's sense of general self-efficacy, which will increase with a deepening use of social media [26]. In summary, the use of social media can improve people's self-efficacy at work in different ways. Based on the above conclusions, we put forward the following hypotheses:

> H3: The use of social media at work can improve older workers' self-efficacy.
>
> H4: The emotional support accessed from social media can improve older workers' self-efficacy.
>
> H5: The information support accessed from social media can improve older workers' self-efficacy.

Social Support and Work Ability

Work ability is defined as the sum of the factors enabling an employed person in a certain situation to manage his or her working demands successfully [27]. Older workers have abundant work experience and good working skills, but they inevitably experience a decline in physical fitness and cognitive ability with an increase in age. Therefore, for older people, work ability refers to the physical and mental health level that can meet the needs of the work [28].

Effective social support is of great significance to the work ability of employees, especially for older workers. Effective emotional support is an important way for people to maintain and improve their working ability. For example, Pohl and Galletta have demonstrated that emotional support provided by a supervisor in the workplace can improve employees' work ability and reduce their sense of fatigue [29]. A study by Karlsson et al showed that emotional support can help individuals better cope with injuries and combat the decline in work ability caused by injuries [30]. In addition, the work ability level is often associated with the mental health level. Especially for older workers, mental illness is an important reason for a decline in work ability [31]. Emotional support can help individuals prevent and alleviate the harm brought by mental illness [17].

Similarly, information support can also have an important impact on an individual's work ability. Information support at work

mainly includes information sharing and the exchange of new knowledge and skills [19]. For older workers, learning new knowledge and skills can improve their work efficiency and prevent the decline in work ability caused by aging [32]. In addition, non–work-related information sharing also can improve older people's ability to work. For example, Edmunds et al confirmed that the support of colleagues related to health information can promote workplace exercise, improve workers' health level, and enhance their work ability [11]. Workplace exercise can not only improve the physical health of older adults but also improve their mental health, which is of great significance for maintaining their work ability [33]. To sum up, we believe that both emotional support and information support will impact the working ability of older workers. Therefore, we hypothesize the following:

H6: Emotional support can improve older workers' work ability.

H7: Information support can improve older workers' work ability.

Work stress reflects an interaction between individual characteristics and an individual's response to work characteristics [34]. Ganster and Rosen argued that work stress is a process where workers experience mental and physical changes in the short or long term caused by mental experiences and demands at the workplace (the source of stress) [35]. Work stress has negative effects on individuals, causing such physical problems as headache, heart disease, elevated blood pressure, gastropathy, and insomnia, and psychological disorders including depression, hostility, and withdrawal [36].

Good social support can improve employee performance and ease work stress [37]. Emotional support can help individuals better cope with work stress. For example, research by Yang et al shows that peer support and subjective emotional support can effectively help older employees cope with work stress [38]. Moeller and Chung-Yan also confirmed that emotional support from supervisors can improve employees' mental health and reduce their work stress [39]. Similarly, information support from supervisors can also effectively help employees cope with work stress [39]. This is because credible workplace information can reduce an employee's sense of unpredictability and powerlessness, which reduces psychological distress and work stress [39]. In addition, research by Chrisopoulos et al suggests that although emotional support can improve people's ability to cope with work stress, it cannot change the objective stress situation; information support related to tasks or technologies can improve people's work efficiency and reduce their work stress [40]. Based on the above conclusions, we put forward the following hypotheses:

H8: Emotional support can relieve older workers' work stress.

H9: Information support can relieve older workers' work stress.

Self-Efficacy and Work Ability

Self-efficacy greatly influences an individual's abilities and work stress [41]. First, individuals with higher self-efficacy can cope with work stress more effectively and are more positive about their work. Skaalvik and Skaalvik found a negative correlation between teacher stress and teacher self-efficacy [42]. Research by Lloyd et al shows that self-efficacy can improve people's intrinsic work motivation and reduce work stress [43]. Self-efficacy can also improve work ability in different dimensions. Self-efficacy can influence people's work ability by improving an individual's health, self-confidence, social function, and other factors [44]. Improving work-related self-efficacy and self-management can improve employees' work ability [45]. Higher self-efficacy can also reduce the incidence of mental illness [46]. A healthy state of mind can have a positive effect on work ability and status [31]. Based on the above findings, we believe that self-efficacy can improve older people's work ability and relieve their work stress. Therefore, we put forward the following hypotheses:

H10: Self-efficacy can improve older workers' work ability.

H11: Self-efficacy can alleviate older workers' work stress.

Willingness to Delay Retirement

The main influencing factors on the duration of older individuals' work lives are their individual health and financial pressures [12]. An individual's subjective will also plays a decisive role [12]. For example, Skaalvik and Skaalvik found that work stress is an important factor affecting one's willingness to retire [42]. Heavier work stress can lead to early retirement, while lighter work stress can prompt individuals to delay their retirement, even motivating teachers older than 65 years to continue working. The individual's sense of satisfaction about nonmaterial factors, such as enjoying working and social recognition, will affect the willingness of adults who reach retirement age to delay retirement [47]. Highly competent (as far as work ability) individuals tend to have greater enthusiasm for work. For individuals with greater enthusiasm for work, retirement can lead to a huge psychological gap [48]. In addition, when work is challenging, people with high work ability tend to be more willing to work longer and delay their retirement [49]. Therefore, for individuals with a passion for work and higher level of work ability, extending the duration of their working life as much as possible can help them maintain their existing social status and enjoy the fun of work. Therefore, we put forward the following hypotheses:

H12: Work stress will have a negative impact on older workers' willingness to delay retirement.

H13: Work ability will have a positive impact on older workers' willingness to delay retirement.

Finally, the social support gained at work also influences an individual's willingness to retire. According to Hofstetter and Cohen, for older adults, emotional support from colleagues (such as kindness and acceptance) will enhance people's happiness and affect their retirement intention [50]. Emotional support provided by an organization (such as caring for workers' general health) will affect the sense of belonging of older workers and their intention to retire [50]. In addition, continuous learning and development (information support) opportunities provided by organizations relate to the preference for postponing

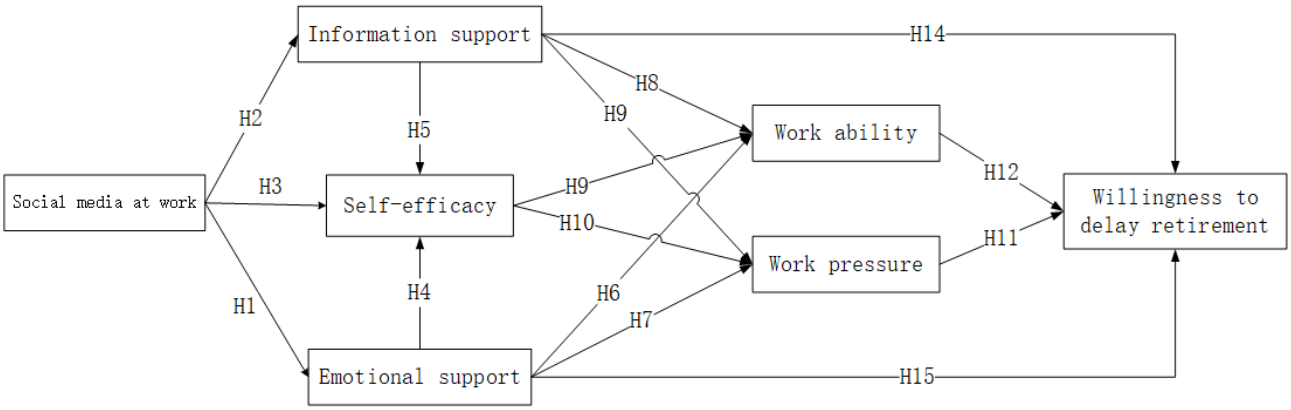
retirement [51]. By contrast, insufficient professional information support by colleagues will lead to career and job stagnation, which may lead to premature retirement [50]. Therefore, the work duration of older individuals is affected by social support. In summary, we believe that work stress, work ability, and social support impact older individuals' willingness to delay retirement. Therefore, we put forward the following hypotheses:

H14: Information support will have a positive impact on older workers' willingness to delay retirement.

H15: Emotional support will have a positive impact on older workers' willingness to delay retirement.

In summary, this study's proposed model studies the willingness of older workers to delay retirement from the perspective of their social media use at work, examining numerous hypotheses as depicted in Figure 1.

Figure 1. The research model.



Methods

Scale Development

In this study, the main measurement variables included social media at work, social support (emotional and information), work ability, willingness to delay retirement, work stress, and self-efficacy. The questionnaire adopted the form of a Likert 7-point scale. An answer from 1 to 7 indicated the degree to which respondents agreed with the question; 1 meant completely disagree, and 7 meant completely agree. All constructs' measures used in this study are listed in Multimedia Appendix 1 [9,52-59].

A scale showing the effects of social media at work was compiled on the basis of previous studies on an effective scale and was amended to adapt to this study's background. The questions to measure included those about frequency, improving work efficiency, improving communication skills, relieving pressure, learning on the job, and undertaking recreational activities [9,52,53]. In the presurvey sample, the scale had good internal consistency reliability, with a Cronbach α of .902.

The scale of work ability and willingness to delay retirement was based on the work ability index [54]. The Chinese version of the scale shows good reliability and validity [55]. We included 7 items and adapted them to fit the current context. The measured items for work ability included the self-evaluation of work ability, mental adaptability to the current work situation, and physical adaptability to the current work situation, as well as a work performance evaluation. The measured items for willingness to delay retirement included the willingness to retire, self-predicting of future work ability, and willingness to delay retirement. In the presurvey sample, the scale had good internal consistency reliability, with a Cronbach α of .873.

To measure social support, we adopted the social support scale by Cohen and Wills [60]. The scale has shown good reliability and validity in previous studies [56,57]. We adapted the original scale to our needs; the measure of dimensionality included information support and emotional support. In the presurvey sample, the scale had good internal consistency reliability, with a Cronbach α of .901.

Work stress was based on a job content questionnaire [61]. The Chinese version of the scale shows good reliability and validity [62]. The scale included 4 items that refer to quantitative, demanding aspects of the job (eg, time pressure, working hard, excessive work). In the presurvey sample, the scale had good internal consistency reliability, with a Cronbach α of .796.

The earliest general self-efficacy scale was compiled by Schwarzerin in 1981, and the Chinese version was compiled and used in 1995 [58]. The scale has shown good reliability and validity in previous studies [58]. We adapted the original scale to our needs and included 4 items in the questionnaire. In the presurvey sample, the scale had good internal consistency reliability, with a Cronbach α of .789.

Data Collection

The data collection process was divided into several stages. To ensure the quality of the questionnaire, 50 presurvey copies were randomly distributed in Bengbu, China; of these, 47 valid questionnaires were recovered. Some participants were interviewed to determine whether there were problems with the questionnaire, such as unclear language expression or rhetorical errors. Based on the presurvey results, we modified the questionnaire.

The formal questionnaire respondents (staff older than 55 years) were obtained through the Bengbu social security office, and they obtained the questionnaire from social media from June to July 2018. A total of 1500 questionnaires were issued, and

1291 were returned, giving a return rate of 86.1%. After eliminating invalid questionnaires (those that had many blank answers and a high repetition rate of answers), 1020 valid questionnaires remained, giving an effective rate of 79.0%. Table 1 shows the statistics for the demographics.

Table 1. The demographics of the sample.

Category	n (%)
Sex	
Male	521 (51.1)
Female	499 (48.9)
Age (years)	
55-60	146 (14.3)
61-65	629 (61.7)
66-70	175 (17.2)
>70	70 (6.9)
Marital status	
Unmarried	61 (5.9)
Married	772 (75.7)
Divorced	76 (7.5)
Widowed	111 (10.9)
Educational background	
Elementary school	328 (32.2)
Middle school	317 (31.1)
High school	106 (10.4)
College	122 (12.0)
Master's degree	147 (14.4)
Income (¥)	
<1000	180 (17.6)
1000-2000	216 (21.2)
2000-3000	210 (20.6)
3000-4000	216 (21.2)
4000-5000	81 (7.9)
5000-6000	64 (6.3)
＞6000	53 (5.2)

Results

Model Overview

We used a structural partial least squares structural equation modeling (PLS-SEM) method to analyze the data obtained. The model framework was analyzed using SmartPLS 3.28 (SmartPLS GmbH). The PLS method has relatively loose requirements for the normal distribution of the research sample data and has flexibility in dealing with missing data. Therefore, PLS is suitable for exploratory factor analysis. In addition, PLS-SEM is a comprehensive method that can simultaneously examine all the relationships between the constructs in the measurement and the structural models and can also handle complex models with direct and indirect relationships [63,64]. Therefore, when the model complexity is high, PLS-SEM has more advantages than other methods.

Measurement Model

The measurement model includes the following steps: First, SPSS 22 (IBM Corporation) was used for data analysis. The Cronbach α coefficient value was .880, greater than .80, indicating that the reliability of the questionnaire was good. The test results show a Kaiser-Meyer-Olkin value of 0.899 and a significance level of $P<.001$. These values indicate that the scale used in this paper has good structural validity, and that the questionnaire is suitable for factor analysis.

To avoid multicollinearity, we tested the data. The maximum variance expansion coefficient was 2.015, much lower than the prescriptive diagnosis of 5 [65]. Moreover, for the goodness of

fit, the standardized root mean square residual was measured. The standardized root mean square residual has already been used as the goodness of fit method in PLS-SEM measurement [66]. Standardized root mean square residual values of less than 0.10 or 0.08 (in the more conservative version) are considered suitable [66]. In this study, the standardized root mean square residual was 0.045, less than 0.08. Thus, the model was very well adapted.

Table 2 shows the statistical data of factor loading, composite reliability, Cronbach α, and average variance extracted (AVE). According to Hair, the value of the Cronbach α coefficient should be above .7; in this study, the Cronbach α coefficient was between .792 and .880, indicating that the reliability of the questionnaire was good [65]. The composite reliability value ranged from 0.873 to 0.909, which was higher than 0.7, indicating that the questionnaire had good convergent validity [65]. In addition, the average variance extracted was greater than 0.5, indicating that the observed items explain the variance more than the error term [65] and that the model aggregation validity is relatively high.

Table 2. Construct reliability and convergent validity.

Construct items	Loading	CR[a]	Cronbach α	AVE[b]
WDR[c]		0.878	.792	0.706
WDR1	0.839	N/A[d]	N/A	N/A
WDR2	0.818	N/A	N/A	N/A
WDR3	0.863	N/A	N/A	N/A
WS[e]		0.891	.838	0.673
WS1	0.798	N/A	N/A	N/A
WS2	0.845	N/A	N/A	N/A
WS3	0.811	N/A	N/A	N/A
WS4	0.825	N/A	N/A	N/A
ES[f]		0.883	.823	0.653
ES1	0.835	N/A	N/A	N/A
ES2	0.821	N/A	N/A	N/A
ES3	0.793	N/A	N/A	N/A
ES4	0.783	N/A	N/A	N/A
SMW[g]		0.909	.880	0.625
SMW1	0.792	N/A	N/A	N/A
SMW2	0.777	N/A	N/A	N/A
SMW3	0.822	N/A	N/A	N/A
SMW4	0.759	N/A	N/A	N/A
SMW5	0.789	N/A	N/A	N/A
SMW6	0.804	N/A	N/A	N/A
IS[h]		0.875	.810	0.636
IST1	0.807	N/A	N/A	N/A
IST2	0.8	N/A	N/A	N/A
IST3	0.8	N/A	N/A	N/A
IST4	0.784	N/A	N/A	N/A
SE[i]		0.893	.840	0.676
SE1	0.814	N/A	N/A	N/A
SE2	0.812	N/A	N/A	N/A
SE3	0.832	N/A	N/A	N/A
SE4	0.831	N/A	N/A	N/A
WAI[j]		0.873	.806	0.632
WAI1	0.78	N/A	N/A	N/A
WAI2	0.802	N/A	N/A	N/A
WAI3	0.8	N/A	N/A	N/A
WAI4	0.799	N/A	N/A	N/A

[a]CR: composite reliability.

[b]AVE: average variance extracted.

[c]WDR: willingness to delay retirement.

[d]N/A: not applicable.

[e]WS: work stress.

[f]ES: emotional support.

[g]SMW: social media at work.

[h]IS: information support.

[i]SE: self-efficacy.

[j]WAI: work ability index.

Table 3 shows that the square root of each factor's AVE value is greater than the other factor correlation coefficients, indicating that the questionnaire had good discriminant validity [65]. In summary, the model has good reliability and validity.

Table 3. Measurement model results.[a]

Constructs	WS[b]	WAI[c]	IS[d]	ES[e]	WDR[f]	SMW[g]	SE[h]
WS	0.820	—[j]	—	—	—	—	—
WAI	0.741	0.795	—	—	—	—	—
IS	0.495	0.510	0.798	—	—	—	—
ES	0.469	0.477	0.460	0.808	—	—	—
WDR	0.703	0.710	0.492	0.479	0.840	—	—
SMW	0.488	0.467	0.456	0.445	0.433	0.791	—
SE	0.481	0.516	0.726	0.412	0.475	0.485	0.822

[a]The numbers on the diagonal are the square roots of the variance shared between the constructs and their measures. Off-diagonal elements are correlations among constructs. For discriminant validity, diagonal elements should be larger than off-diagonal elements.

[b]WS: work stress.

[c]WAI: work ability index.

[d]IS: information support.

[e]ES: emotional support.

[f]WDR: willingness to delay retirement.

[g]SMW: social media at work.

[h]SE: self-efficacy.

[j]—: not applicable.

Structural Model

We used SmartPLS 3.28 to calculate the significance of the model path coefficients using bootstrapping with 3000 samples and 2-tailed *t* tests. The results of direct effects are shown in Figure 2 and Table 4. We also test the total and indirect effects of social media at work on willingness to delay retirement, and the results are shown in Multimedia Appendix 2.

Figure 2. Model results. *P<.05, ***P<.001.

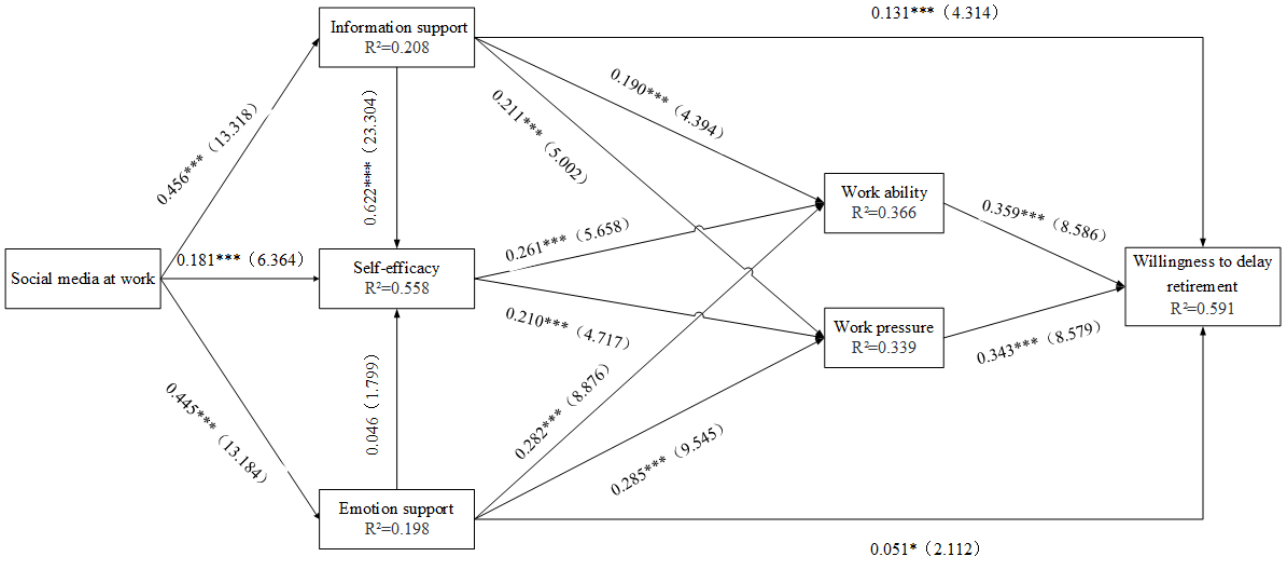


Table 4. Structural parameter estimate.

Hypothesized path	SPC[a]	*t* value	*P* value	Results
H1: The use of social media at work can improve emotional support for older people.	0.445	13.184	<.001	Supported
H2: The use of social media at work can improve information support for older people.	0.456	13.318	<.001	Supported
H3: The use of social media at work can improve older people's self-efficacy.	0.181	6.364	<.001	Supported
H4: The emotional support accessed from social media can improve older people's self-efficacy.	0.046	1.799	.07	Not supported
H5: The information support accessed from social media can improve older people's self-efficacy.	0.622	23.304	<.001	Supported
H6: Emotional support can improve older workers' work ability.	0.282	8.876	<.001	Supported
H7: Information support can improve older workers' work ability.	0.19	4.394	<.001	Supported
H8: Emotional support can relieve older workers' work stress.	0.285	9.545	<.001	Supported
H9: Information support can relieve older workers' work stress.	0.211	5.002	<.001	Supported
H10: Self-efficacy can improve older workers' work ability.	0.261	5.658	<.001	Supported
H11: Self-efficacy can alleviate older workers' work stress.	0.210	4.717	<.001	Supported
H12: Work stress will have a negative impact on older workers' willingness to delay retirement.	0.343	8.579	<.001	Supported
H13: Work ability will have a positive impact on older workers' willingness to delay retirement.	0.359	8.586	<.001	Supported
H14: Information support will have a positive impact on older workers' willingness to delay retirement.	0.131	4.314	<.001	Supported
H15: Emotional support will have a positive impact on older workers' willingness to delay retirement.	0.051	2.112	.04	Supported

[a]SPC: standardized path coefficient.

The hypotheses (H1, H2, and H3) that use of social media at work can improve social support for older people are supported. The effect levels of emotional support (standardized path coefficient=0.445, t_{14}=13.184, P<.001) and of information support (standardized path coefficient=0.456, t_{14}=13.318, P<.001), which are the two dimensions of social support, are quite similar. Compared with social support, social media has a slighter impact on self-efficacy (standardized path coefficient=0.181, t_{14}=6.364, P<.001).

In the hypothesis (H5) of the relationship between social support and self-efficacy, the influence of information support on self-efficacy is confirmed (standardized path coefficient=0.622, t_{14}=23.304, P<.001). However, the hypothesis (H4) on the effect of emotional support on self-efficacy proved to be untenable (standardized path coefficient=0.046, t_{14}=1.799, P=0.07).

The hypotheses that social support has a positive effect on the work ability (H6, H7) and work stress (H8, H9) of older workers are confirmed. The influences of emotional support on the work stress of older adults (standardized path coefficient=0.285, t_{14}=9.545, P<.001) and the ability to work (standardized path coefficient=0.282, t_{14}=8.876, P<.001) are bigger overall than those of information support for older workers' work ability (standardized path coefficient=0.190, t_{14}=4.394, P<.001) and work stress (standardized path coefficient=0.211, t_{14}=5.002, P<.001).

The hypotheses that self-efficacy has a positive effect on the working ability (H10; standardized path coefficient=0.261, t_{14}=5.658, P<.001) and work stress (H11; standardized path coefficient=0.210, t_{14}=4.717, P<.001) of older workers are confirmed.

The hypotheses that work stress (H12), work ability (H13), and social support (H14, H15) have an impact on the willingness of older workers to delay retirement are confirmed. Work stress (standardized path coefficient=0.343, t_{14}=8.579, P<.001) and work ability (standardized path coefficient=0.359, t_{14}=8.586, P<.001) have a greater impact on the retirement intention of older workers than does social support. The influence of emotional support (standardized path coefficient=0.051, t_{14}=2.112, P=.04) on the intention to delay retirement, which is included in social support, is smaller than that of information support (standardized path coefficient=0.131, t_{14}=4.314, P<.001).

Finally, we analyze the effect of social media use at work on older workers' willingness to retire after adding control variables. The *t* test results show that sex, age, marital status, educational level, and income have no significant effect on older workers' willingness to retire. This means that demographic characteristics have no significant effect on the analysis results.

Discussion

Findings

This paper examines the impact of social media use at work on elderly workers' willingness to retire. The empirical results show that 14 of the 15 hypotheses in the research model are confirmed, and 1 hypothesis is not supported. Social media positively impacts elderly workers' willingness to delay retirement. Information support affects their self-efficacy more than emotional support does, while emotional support has a greater effect on work ability and work stress. However, self-efficacy also impacts work ability and the ability to regulate work stress. The impact of work ability and work stress on

willingness to retire is greater than that of social support. More detailed results are given in Multimedia Appendix 2.

Our study confirms that the use of social media at work has a positive impact on older workers. The findings suggest that the use of social media at work can help older workers improve their social support and self-efficacy (H1, H2, and H3), which is very important for older people; effective social support can enhance their physical and mental health and reduce the occurrence of psychological disorders such as depression and anxiety [11,17,33]. Similarly, self-efficacy also has positive implications for the mental health of older people [44,46].

In a work environment, information support is the main factor affecting the self-efficacy of older workers, as compared with emotional support. This is a new finding. In previous studies, there was a significant correlation between social support and self-efficacy [23-25]. However, in this study, only information support impacts self-efficacy (H5), and emotional support has no influence on self-efficacy (H4). This result can be explained by the fact that for older workers, past work experience helps them build enough confidence to complete the work they are given. Therefore, emotional encouragement and support have no significant effect on their work self-efficacy. In addition, information support can help older people understand relevant information and learn new skills needed in their work. This is a major finding that reveals new ways to improve the self-efficacy of older workers.

The results also show that the impact of emotional support on work ability and work stress is generally higher than that of information support (H6-H9). This result can be explained by the fact that older workers have established practices for handling their work based on past experience, and the acquisition of new knowledge and new skills is only complementary to their own work ability. They are also experienced enough to cope with work stress. For older workers, who are moving toward the last years of their lives, effective emotional support such as a sense of achievement and satisfaction, as well as respect, can help them to work better and handle work stress. Therefore, the impact of information support on older workers is less than that of emotional support, in this context.

The results show that self-efficacy impacts work ability and ability to regulate work stress (H10 and H11). This result confirms the results of previous research [41]. We believe that this result is due to the specific life stage of older adults. The intellectual and physical decline caused by aging requires effective internal motivation to help older adults stay active at work and alleviate work stress. In addition, considering the results of H6 through H9, social support can alleviate work stress. We found that compared with self-efficacy, social support can more effectively alleviate work stress for older workers. Considering the sources of work stress, we believe that for older workers, when they are doing complex work, they are less able to regulate emotions internally, and external support is needed to help them moderate their emotions. This also confirms the research of Isaacowitz et al [67].

The impact of work ability and work stress on willingness to delay retirement is greater than that of social support. This result indicates that among the influencing factors for older workers' willingness to delay retirement, the individual perceptions of work ability and work stress are the main factors. In addition, work ability and work stress have an equally important impact on one's willingness to delay retirement, while social support has a relatively small impact on this willingness. This result suggests that older people's willingness to delay retirement is mainly affected by work-related factors, and stronger work ability and lower work stress can extend their work duration and delay their retirement.

Implications for Research

This study has several theoretical contributions.

First, our study links the use of social media at work to the willingness of older workers to delay retirement. This is a topic that has not been considered in the past. Past research on the use of social media at work was not specifically targeted at older workers, and there was limited research on their willingness to delay retirement. This study fills this gap. It also demonstrates the influencing mechanism of social media on the special group of older workers. This finding reveals the positive role played by the use of social media at work and its applicability across different groups.

In the course of the study, we explored the impact of social support, work ability, and work stress on older people's willingness to delay retirement, which has never been explored before. The results show that an individual's work ability (internal factor) and work stress (external factor) are mainly affected by external support, while internal subjective motivation (self-efficacy) plays a smaller role than external support. This finding can provide a theoretical basis for guiding the establishment of an effective incentive model for delayed retirement in the future.

Second, our study shows the special nature of older workers. The link between social support and self-efficacy has long been confirmed via past studies [23]. It is generally argued that all dimensions of social support have a positive impact on self-efficacy. However, in this study, the emotional support dimension of social support does not influence self-efficacy. This result suggests that in the working environment, the social support that older workers gain from using social media at work is more intentional; that is, the support is in getting help, gaining new knowledge, or reducing stress at work. This result confirms the results of previous studies [9].

At the same time, we also found that emotional support at work is more important than information support for older people in terms of work ability and work stress. This result shows the special nature of older workers. That is, as older workers are at a late stage of life, they need emotional support more than other dimensions of social support. This result also confirms that different dimensions of social support play different roles at different stages. This is an interesting finding that develops and extends social support–related theories.

Research has shown that self-efficacy is often an important factor to help individuals cope with work stress and improve work ability. This result also confirms the results of previous studies [42]. Combined with the impact of social support on

self-efficacy (H4 and H5), the result shows that the self-efficacy of older workers often comes from their acquisition of technology experience and ability, while successful practices from their pasts weaken the role of external emotional incentives. Considering the differences between this study and previous studies (age differences, work and nonwork differences), the results of this study help to further understanding of the special nature of older workers.

Implications for Practice

This study contributes to practice in the following ways:

Our study can provide a basis for relevant government departments in their development of a health bonus plan. Especially in China, the rapid aging and fertility decline of the overall population create an urgent need to develop a health bonus plan in line with national conditions. Moreover, our study can reduce resistance from older people when developing a deferred-retirement policy. The results of the study show that older people's willingness to delay retirement is influenced by their work ability and work stress. Therefore, the differences between occupations and the actual situation of different types of groups of older adults can be taken into account during the formulation of a deferred-retirement plan, so as to make the plan more reasonable and effectively use the resources of the older adult population and promote the realization of the health bonus.

For companies, our study helps improve the management of, and optimize management plans for, older employees. Based on our conclusions, emotional support can better improve work ability and can reduce work stress more effectively for older workers. Therefore, companies can provide older employees with more emotional encouragement and care and can pay attention to their emotional state. From a social media perspective, companies should encourage and support older people in their use of social media at work, which is of positive significance. For example, company executives can show concern for older workers regarding their work status and mood by using social media. Additionally, colleagues can exchange work-related information, such as new knowledge and new job skills, or encourage each other with exercising by using social media.

The older workers' acceptance and use of social media devices is important to realize positive aging; this is also an important prerequisite for realizing the health bonus. Our study confirms that using social media at work can improve both social support and self-efficacy, which can not only help older adults at work but also improve their mental health to prevent the physical and mental health problems caused by aging. This result can also help social media developers improve and develop social media features that are more suitable for older people.

Limitations and Future Directions

Due to limited time and other constraints, our study has the following limitations:

First, we studied the impact of social media on older people's willingness to retire. While this impact is confirmed, social support accessed from social media is only studied in terms of information support and emotional support. Social support has multiple dimensions, and our study does not explore the impact of these other dimensions of social support on older people; examples include objective social support and perceived social support dimensions, and the impact of the availability of support on future deferred-retirement intentions. Therefore, future studies will subdivide social support into these other dimensions and explore in detail the relationship between different social support dimensions, in order to understand the intrinsic relevance of social support to older people's willingness to delay retirement.

Second, this study does not subdivide occupations or take into account the different characteristics of different occupations. Our study is incomplete, and in the future different occupations will be compared to explore the differences in the willingness to delay retirement among older people in different occupations.

Finally, the main object of study in this paper is a segment of the older population in Anhui Province, China. It is unclear whether the same results would be obtained in other provinces or countries due to cultural differences. Future research based on the results of this study will compare the willingness to delay retirement in different cultures.

Conclusions

The need for deferred retirement has gained a general consensus in China and other developed countries. As China's population is experiencing an accelerating aging process, how to promote the transformation of the demographic bonus to a health bonus and effectively formulate a deferred-retirement policy has become an urgent problem in China. This paper examines factors influencing older people's willingness to delay retirement from the perspective of social media. The results of the study provide relevant references for solving this problem.

Our study shows that for older workers, the willingness to delay retirement is mainly affected by work ability and external work stress. Social support gained from social media can effectively help older people enhance their work ability and ease work stress. The results of the model hypothesis test provide the characteristics of older workers' need of social support.

Based on the findings of this paper, we suggest that the government create deferred-retirement plans based on different occupations and demographic characteristics. We have put forward proposals to extend people's working lives and help governments implement health bonus policies. Older adults themselves can also actively use social media to improve their social support and physical and psychological health.

Acknowledgments

The data set collection and analysis of this research were partially funded by the National Natural Science Foundation of China, grant numbers 71771075, 71771077 and 72071063, and Fundamental Research Funds for the central universities, grant number PA2020GDKC0020.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Measurement scale.
[DOCX File , 22 KB - jmir_v23i2e18264_app1.docx]

Multimedia Appendix 2
Indirect and total effects of social media at work on willingness to delay retirement.
[DOCX File , 16 KB - jmir_v23i2e18264_app2.docx]

References

1. WHO. Mental health of older adults. Geneva: World Health Organization; 2017. URL: http://www.who.int/en/news-room/fact-sheets/detail/mental-health-of-older-adults [accessed 2021-01-15]
2. Bloom DE, Canning D, Fink G, Finlay JE. Fertility, female labor force participation, and the demographic dividend. J Econ Growth 2009 May 5;14(2):79-101. [doi: 10.1007/s10887-009-9039-9]
3. WHO. World health statistics 2016: monitoring health for the SDGs sustainable development goals. Geneva: World Health Organization; 2016.
4. Mason A, Lee R. Reform and support systems for the elderly in developing countries: capturing the second demographic dividend. Genus 2006;62(2):11-35.
5. Hongchun W. From Demographic Bonus to Health Bonus. Medicine & Philosophy 2006(05):21-22 [FREE Full text]
6. Ballesteros S, Mayas J, Ruiz-Marquez E, Prieto A, Toril P, Ponce de Leon L, et al. Effects of Video Game Training on Behavioral and Electrophysiological Measures of Attention and Memory: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2017 Jan 24;6(1):e8 [FREE Full text] [doi: 10.2196/resprot.6570] [Medline: 28119279]
7. Mattsson S, Olsson EMG, Alfonsson S, Johansson B, Carlsson M. Measuring Use of Health-Related Support on the Internet: Development of the Health Online Support Questionnaire (HOSQ). J Med Internet Res 2015 Nov 20;17(11):e266 [FREE Full text] [doi: 10.2196/jmir.4425] [Medline: 26589638]
8. Wang P, Chaudhry S, Li L. Exploring the influence of social media on employee work performance. Internet Research Apr 2016;26(2):529-545. [doi: 10.1108/intr-11-2014-0299]
9. Leftheriotis I, Giannakos MN. Using social media for work: Losing your time or improving your work? Computers in Human Behavior 2014 Feb;31:134-142. [doi: 10.1016/j.chb.2013.10.016]
10. Kwahk K, Park D. The effects of network sharing on knowledge-sharing activities and job performance in enterprise social media environments. Computers in Human Behavior 2016 Feb;55:826-839. [doi: 10.1016/j.chb.2015.09.044]
11. Edmunds S, Sitch M, Lowry R. Who provides physical activity support in the workplace? Implications for peer led interventions. Health Education Journal 2019 Sep 12;79(2):195-211. [doi: 10.1177/0017896919872230]
12. Wang M, Shultz KS. Employee Retirement: A Review and Recommendations for Future Investigation. Journal of Management 2009 Oct 06;36(1):172-206. [doi: 10.1177/0149206309347957]
13. Norheim KL, Hjort Bønløkke J, Samani A, Omland Ø, Madeleine P. The Effect of Aging on Physical Performance Among Elderly Manual Workers: Protocol of a Cross-Sectional Study. JMIR Res Protoc 2017 Nov 22;6(11):e226 [FREE Full text] [doi: 10.2196/resprot.8196] [Medline: 29167091]
14. Kietzmann JH, Hermkens K, McCarthy IP, Silvestre BS. Social media? Get serious! Understanding the functional building blocks of social media. Business Horizons 2011 May;54(3):241-251. [doi: 10.1016/j.bushor.2011.01.005]
15. El Ouirdi A, El Ouirdi M, Segers J, Henderickx E. Employees' use of social media technologies: a methodological and thematic review. Behaviour & Information Technology 2015 Jan 22;34(5):454-464. [doi: 10.1080/0144929x.2015.1004647]
16. Cetinkaya, AS, Rashid M. The Effect of Social Media on Employees' Job Performance: The mediating Role of Organizational Structure. Journal of Organizational Psychology 2018 Nov 1;18(4):94-116. [doi: 10.33423/jop.v18i4.88]
17. Gottlieb B. Social support strategies: Guidelines for mental health practice. Los Angeles: SAGE Publications, Incorporated; 1983.
18. Charoensukmongkol P. Effects of support and job demands on social media use and work outcomes. Computers in Human Behavior 2014 Jul;36:340-349. [doi: 10.1016/j.chb.2014.03.061]

19. Skeels MM, Grudin J. When social networks cross boundaries: a case study of workplace use of facebook and linkedin. 2009 Presented at: Proceedings of the ACM international conference on Supporting group work; 2009; New York. [doi: 10.1145/1531674.1531689]
20. Lewis ZH, Swartz MC, Martinez E, Lyons EJ. Social Support Patterns of Middle-Aged and Older Adults Within a Physical Activity App: Secondary Mixed Method Analysis. JMIR Aging 2019 Aug 23;2(2):e12496 [FREE Full text] [doi: 10.2196/12496] [Medline: 31518281]
21. Olaya B, Domènech-Abella J, Moneta MV, Lara E, Caballero FF, Rico-Uribe LA, et al. All-cause mortality and multimorbidity in older adults: The role of social support and loneliness. Exp Gerontol 2017 Dec 01;99:120-126. [doi: 10.1016/j.exger.2017.10.001] [Medline: 28982608]
22. Bandura A, Adams NE. Analysis of self-efficacy theory of behavioral change. Cogn Ther Res 1977 Dec;1(4):287-310. [doi: 10.1007/BF01663995]
23. Nwoke MB, Onuigbo EN, Odo VO. Social support, self-efficacy and gender as predictors of reported stress among inpatient caregivers. The Social Science Journal 2019 Dec 09;54(2):115-119. [doi: 10.1016/j.soscij.2016.08.001]
24. Tims M, Bakker AB, Xanthopoulou D. Do transformational leaders enhance their followers' daily work engagement? The Leadership Quarterly 2011 Feb;22(1):121-131. [doi: 10.1016/j.leaqua.2010.12.011]
25. Xanthopoulou D, Bakker AB, Heuven E, Demerouti E, Schaufeli WB. Working in the sky: a diary study on work engagement among flight attendants. J Occup Health Psychol 2008 Oct;13(4):345-356. [doi: 10.1037/1076-8998.13.4.345] [Medline: 18837629]
26. Bright LF, Kleiser SB, Grau SL. Too much Facebook? An exploratory examination of social media fatigue. Computers in Human Behavior 2015 Mar;44:148-155. [doi: 10.1016/j.chb.2014.11.048]
27. Mache S, Vitzthum K, Groneberg DA. High work ability and performance in psychiatric health care services: Associations with organizational, individual and contextual factors. Neurology, Psychiatry and Brain Research 2015 Jun;21(2):88-95. [doi: 10.1016/j.npbr.2015.03.001]
28. Ilmarinen J. Work ability--a comprehensive concept for occupational health research and prevention. Scand J Work Environ Health 2009 Jan;35(1):1-5 [FREE Full text] [doi: 10.5271/sjweh.1304] [Medline: 19277432]
29. Pohl S, Galletta M. The role of supervisor emotional support on individual job satisfaction: A multilevel analysis. Appl Nurs Res 2017 Feb;33:61-66. [doi: 10.1016/j.apnr.2016.10.004] [Medline: 28096025]
30. Karlsson N, Skargren E, Kristenson M. Emotional support predicts more sickness absence and poorer self assessed work ability: a two-year prospective cohort study. BMC Public Health 2010 Oct 26;10:648 [FREE Full text] [doi: 10.1186/1471-2458-10-648] [Medline: 20977767]
31. Karpov B, Joffe G, Aaltonen K, Suvisaari J, Baryshnikov I, Näätänen P, et al. Level of functioning, perceived work ability, and work status among psychiatric patients with major mental disorders. Eur Psychiatry 2017 Jul;44:83-89. [doi: 10.1016/j.eurpsy.2017.03.010] [Medline: 28545013]
32. Tuomi K, Huuhtanen P, Nykyri E, Ilmarinen J. Promotion of work ability, the quality of work and retirement. Occup Med (Lond) 2001 Aug;51(5):318-324. [doi: 10.1093/occmed/51.5.318] [Medline: 11473138]
33. Poscia A, Moscato U, La Milia DI, Milovanovic S, Stojanovic J, Borghini A, et al. Workplace health promotion for older workers: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16(S5):329 [FREE Full text] [doi: 10.1186/s12913-016-1518-z] [Medline: 27609070]
34. Brunborg GS. Core self-evaluations: A predictor variable for job stress. European Psychologist 2008 Jan;13(2):96-102. [doi: 10.1027/1016-9040.13.2.96]
35. Ganster DC, Rosen CC. Work stress and employee health: A multidisciplinary review. Journal of Management 2013 Feb 19;39(5):1085-1122. [doi: 10.1177/0149206313475815]
36. Gupta N, Beehr TA. Job stress and employee behaviors. Organ Behav Hum Perform 1979 Jun;23(3):373-387. [doi: 10.1016/0030-5073(79)90004-7] [Medline: 10242304]
37. Beehr TA, Jex SM, Stacy BA, Murray MA. Work stressors and coworker support as predictors of individual strain and job performance. J Organiz Behav 2000 Jun;21(4):391-405. [doi: 10.1002/(sici)1099-1379(200006)21:4<391::aid-job15>3.0.co;2-9]
38. Yang T, Shen Y, Zhu M, Liu Y, Deng J, Chen Q, et al. Effects of Co-Worker and Supervisor Support on Job Stress and Presenteeism in an Aging Workforce: A Structural Equation Modelling Approach. Int J Environ Res Public Health 2015 Dec 23;13(1):ijerph13010072 [FREE Full text] [doi: 10.3390/ijerph13010072] [Medline: 26703705]
39. Moeller C, Chung‐Yan GA. Effects of social support on professors' work stress. Intl J of Educational Mgt 2013 Mar 22;27(3):188-202. [doi: 10.1108/09513541311306431]
40. Chrisopoulos S, Dollard M, Winefield A, Dormann C. Increasing the probability of finding an interaction in work stress research: A two‐wave longitudinal test of the triple‐match principle. Journal of Occupational Organizational Psychology 2010;83(1):17-37. [doi: 10.1348/096317909x474173]
41. Stajkovic AD, Luthans F. Self-efficacy and work-related performance: A meta-analysis. Psychological Bulletin 1998;124(2):240-261. [doi: 10.1037/0033-2909.124.2.240]
42. Skaalvik E, Skaalvik S. Teacher Stress and Teacher Self-Efficacy: Relations and Consequences. In: Educator Stress. Switzerland: Springer; 2017:101-125.

43. Lloyd J, Bond FW, Flaxman PE. Work-related self-efficacy as a moderator of the impact of a worksite stress management training intervention: Intrinsic work motivation as a higher order condition of effect. J Occup Health Psychol 2017 Jan;22(1):115-127. [doi: 10.1037/ocp0000026] [Medline: 27054501]
44. Bandura A, Locke EA. Negative self-efficacy and goal effects revisited. J Appl Psychol 2003 Feb;88(1):87-99. [doi: 10.1037/0021-9010.88.1.87] [Medline: 12675397]
45. Linden M, Muschalla B, Hansmeier T, Sandner G. Reduction of sickness absence by an occupational health care management program focusing on self-efficacy and self-management. Work 2014;47(4):485-489. [doi: 10.3233/WOR-131616] [Medline: 23531574]
46. Indregard AR, Knardahl S, Nielsen MB. Emotional Dissonance, Mental Health Complaints, and Sickness Absence Among Health- and Social Workers. The Moderating Role of Self-Efficacy. Front Psychol 2018;9:592 [FREE Full text] [doi: 10.3389/fpsyg.2018.00592] [Medline: 29740375]
47. Hofäcker D, Naumann E. The emerging trend of work beyond retirement age in Germany. Increasing social inequality? Z Gerontol Geriatr 2015 Jul;48(5):473-479. [doi: 10.1007/s00391-014-0669-y] [Medline: 25117859]
48. Houlfort N, Fernet C, Vallerand RJ, Laframboise A, Guay F, Koestner R. The role of passion for work and need satisfaction in psychological adjustment to retirement. Journal of Vocational Behavior 2015 Jun;88:84-94. [doi: 10.1016/j.jvb.2015.02.005]
49. Damman M, Henkens K, Kalmijn M. Women's Retirement Intentions and Behavior: The Role of Childbearing and Marital Histories. Eur J Population 2015 Aug 29;31(4):339-363. [doi: 10.1007/s10680-014-9335-8]
50. Hofstetter H, Cohen A. The mediating role of job content plateau on the relationship between work experience characteristics and early retirement and turnover intentions. Personnel Review 2014;43(3):350-376. [doi: 10.1108/pr-03-2012-0054]
51. Herrbach O, Mignonac K, Vandenberghe C, Negrini A. Perceived HRM practices, organizational commitment, and voluntary early retirement among late-career managers. Hum Resour Manage 2009 Nov;48(6):895-915. [doi: 10.1002/hrm.20321]
52. Tulu DT. Should online social Medias (OSMs) be banned at work? The impact of social Medias on employee productivity in Ambo University, a case study. Research in International Business and Finance 2017 Dec;42:1096-1102. [doi: 10.1016/j.ribaf.2017.07.044]
53. Demircioglu MA, Chen C. Public employees' use of social media: Its impact on need satisfaction and intrinsic work motivation. Government Information Quarterly 2019 Jan;36(1):51-60. [doi: 10.1016/j.giq.2018.11.008]
54. Martus P, Jakob O, Rose U, Seibt R, Freude G. A comparative analysis of the Work Ability Index. Occup Med (Lond) 2010 Oct;60(7):517-524. [doi: 10.1093/occmed/kqq093] [Medline: 20573840]
55. Li H, Liu Z, Liu R, Li L, Lin A. The relationship between work stress and work ability among power supply workers in Guangdong, China: a cross-sectional study. BMC Public Health 2016 Feb 06;16:123 [FREE Full text] [doi: 10.1186/s12889-016-2800-z] [Medline: 26850629]
56. Hobman E, Restubog S, Bordia P, Tang R. Abusive supervision in advising relationships: Investigating the role of social support. Applied Psychology 2009;58(2):233-256. [doi: 10.1111/j.1464-0597.2008.00330.x]
57. Ju C, Lan J, Li Y, Feng W, You X. The mediating role of workplace social support on the relationship between trait emotional intelligence and teacher burnout. Teaching and Teacher Education 2015 Oct;51:58-67. [doi: 10.1016/j.tate.2015.06.001]
58. Cheung S, Sun SYK. Assessment of optimistic self-beliefs: further validation of the Chinese version of the General Self-Efficacy Scale. Psychol Rep 1999 Dec;85(3 Pt 2):1221-1224. [doi: 10.2466/pr0.1999.85.3f.1221] [Medline: 10710976]
59. Cheng Y, Luh W, Guo Y. Reliability and validity of the Chinese version of the Job Content Questionnaire in Taiwanese workers. Int J Behav Med 2003;10(1):15-30. [doi: 10.1207/s15327558ijbm1001_02] [Medline: 12581945]
60. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]
61. Karasek R, Brisson C, Kawakami N, Houtman I, Bongers P, Amick B. The Job Content Questionnaire (JCQ): an instrument for internationally comparative assessments of psychosocial job characteristics. J Occup Health Psychol 1998 Oct;3(4):322-355. [doi: 10.1037//1076-8998.3.4.322] [Medline: 9805280]
62. Li J, Yang W, Liu P, Xu Z, Cho SI. Psychometric evaluation of the Chinese (mainland) version of Job Content Questionnaire: a study in university hospitals. Ind Health 2004 Apr;42(2):260-267 [FREE Full text] [doi: 10.2486/indhealth.42.260] [Medline: 15128178]
63. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. EBR 2019 Jan 14;31(1):2-24. [doi: 10.1108/ebr-11-2018-0203]
64. Henseler J, Ringle C, Sinkovics R. The use of partial least squares path modeling in international marketing. Bingley, UK: Emerald Group Publishing Limited; 2009.
65. Hair JF, Sarstedt M, Ringle CM, Mena JA. An assessment of the use of partial least squares structural equation modeling in marketing research. J of the Acad Mark Sci 2011 Jun 7;40(3):414-433. [doi: 10.1007/s11747-011-0261-6]
66. Rönkkö M, Evermann J. A Critical Examination of Common Beliefs About Partial Least Squares Path Modeling. Organizational Research Methods 2013 Mar 07;16(3):425-448. [doi: 10.1177/1094428112474693]
67. Isaacowitz DM, Livingstone KM, Castro VL. Aging and emotions: experience, regulation, and perception. Curr Opin Psychol 2017 Oct;17:79-83 [FREE Full text] [doi: 10.1016/j.copsyc.2017.06.013] [Medline: 28950978]

Abbreviations

AVE: average variance extracted
PLS-SEM: partial least squares structural equation modeling

Edited by R Kukafka; submitted 19.02.20; peer-reviewed by Z Xi, S Deng, D Frohlich, L Shen, MB Sampa; comments to author 31.08.20; revised version received 25.10.20; accepted 17.12.20; published 10.02.21.

Please cite as:
Ma Y, Liang C, Gu D, Zhao S, Yang X, Wang X
How Social Media Use at Work Affects Improvement of Older People's Willingness to Delay Retirement During Transfer From Demographic Bonus to Health Bonus: Causal Relationship Empirical Study
J Med Internet Res 2021;23(2):e18264
URL: https://www.jmir.org/2021/2/e18264
doi:10.2196/18264
PMID:33565983

Original Paper

Technology-Enabled Reform in a Nontraditional Mental Health Service for Eating Disorders: Participatory Design Study

Alyssa Clare Milton[1], BSc, MAppSc, PhD; Ashlea Hambleton[2], BPsych, MClinPsych; Mitchell Dowling[1], BPsych, PhD; Anna Elizabeth Roberts[1], BAppSC, MExPhys, MBMSc; Tracey Davenport[1], BA, EMBA; Ian Hickie[1], AM, MD, FRANZCP, FASSA

[1]University of Sydney, Sydney, Australia

[2]Innowell, Sydney, Australia

Corresponding Author:
Tracey Davenport, BA, EMBA
University of Sydney
Rm 04.12, Brain and Mind Centre
Sydney, 2050
Australia
Phone: 61 0404839897
Email: Tracey.Davenport@sydney.edu.au

Abstract

Background: The recent Australian National Agenda for Eating Disorders highlights the role technology can play in improving accessibility and service development through web-based prevention, early access pathways, self-help, and recovery assistance. However, engagement with the eating disorders community to co-design, build, and evaluate these much-needed technology solutions through participatory design processes has been lacking and, until recently, underresourced.

Objective: This study aims to customize and configure a technology solution for a nontraditional (web-based, phone, email) mental health service that provides support for eating disorders and body image issues through the use of participatory design processes.

Methods: Participants were recruited chiefly through the Butterfly National Helpline 1800 ED HOPE (Butterfly's National Helpline), an Australian-wide helpline supporting anyone concerned by an eating disorder or body image issue. Participants included individuals with lived experience of eating disorders and body image issues, their supportive others (such as family, health professionals, support workers), and staff of the Butterfly Foundation. Participants took part in participatory design workshops, running up to four hours, which were held nationally in urban and regional locations. The workshop agenda followed an established process of discovery, evaluation, and prototyping. Workshop activities included open and prompted discussion, reviewing working prototypes, creating descriptive artifacts, and developing user journeys. Workshop artifacts were used in a knowledge translation process, which identified key learnings to inform user journeys, user personas, and the customization and configuration of the InnoWell Platform for Butterfly's National Helpline. Further, key themes were identified using thematic techniques and coded in NVivo 12 software.

Results: Six participatory design workshops were held, of which 45 participants took part. Participants highlighted that there is a critical need to address some of the barriers to care, particularly in regional and rural areas. The workshops highlighted seven overarching qualitative themes: identified barriers to care within the current system; need for people to be able to access the right care anywhere, anytime; recommendations for the technological solution (ie, InnoWell Platform features and functionality); need for communication, coordination, and integration of a technological solution embedded in Butterfly's National Helpline; need to consider engagement and tone within the technological solution; identified challenges and areas to consider when implementing a technological solution in the Helpline; and potential outcomes of the technological solution embedded in the Helpline relating to system and service reform. Ultimately, this technology solution should ensure that the right care is provided to individuals the first time.

Conclusions: Our findings highlight the value of actively engaging stakeholders in participatory design processes for the customization and configuration of new technologies. End users can highlight the critical areas of need, which can be used as a catalyst for reform through the implementation of these technologies in nontraditional services.

(J Med Internet Res 2021;23(2):e19532) doi:10.2196/19532

KEYWORDS

eating disorders; body image; mental health; technology; co-design; participatory design; service reform; consumer engagement

Introduction

Over 1 million Australians currently have a clinically diagnosed eating disorder; however, only 25% of Australians with an eating disorder are known to the health system [1]. There has been a recent increase in government funding for the prevention, detection, assessment, and treatment of eating disorders, with a specific focus on using technology solutions to prevent the onset of eating disorders [2]. Across Australia in the last 12 months alone, there has been an Aus $200 million (US $155 million) investment in eating disorders. This includes Aus $70 million (US $54 million) for the establishment of seven residential eating disorder centers around Australia, Aus $4 million (US $3 million) for research translation, Aus $110 million (US $85 million) to fund dedicated Medicare services for eating disorders, and Aus $3 million (US $2 million) to Butterfly National Helpline 1800 ED HOPE (Butterfly's National Helpline).

The Butterfly Foundation released a National Agenda for Eating Disorders with the aim of establishing a baseline for accessible evidence-based treatments for everyone affected by eating disorders in Australia [1]. Within this agenda, technology is identified as a facilitator of improving accessibility and national service development. The agenda identified the importance of developing existing digital services, such as Butterfly's National Helpline to include web-based prevention, early access pathways, self-help, and recovery assistance. Further, technology solutions can address gaps in the continuum of care by extending existing eating disorder-specific web-based and telephone services to include self-help programmes for people with mild or subclinical bulimia nervosa and binge eating disorder and to provide recovery support services for individuals, carers, families, and friends.

In July 2018, The University of Sydney's Brain and Mind Centre and the Butterfly Foundation partnered to engage the client base (individuals, supportive others [SO; eg, Carers, families, and friends] and other professionals affected by an eating disorder, disordered eating, or body image issue) and staff (counselors and service managers) of Butterfly's National Helpline in participatory design workshops to explore how a web-based platform (ie, the InnoWell Platform) could be tailored to realize a technology solution for a nontraditional mental health service, such as Butterfly's National Helpline. In brief, the Helpline is a free and confidential service that provides information, counseling, and treatment referral for eating disorders, disordered eating, body image, and related issues via telephone, web-based chat, or email. Butterfly's National Helpline counselors are professionally trained and experienced in supporting those affected by an eating disorder—the individual who is struggling with their journey and their SO. The Helpline provides information, support, and guidance on treatment options as well as referral pathways on an as-needs basis, and is delivered as a brief intervention with no ongoing therapeutic engagement with those who contact the service. Commonly, the Helpline is the first contact point for an individual with concerns about their eating and body image and those that have not experienced any type of treatment to date [3].

Prior to engagement with the Butterfly Foundation, a partnership project called Project Synergy (Phase 1: 2014-2016) was undertaken by the Young and Well Cooperative Research Centre and The University of Sydney's Brain and Mind Centre. Project Synergy (Phase 1: 2014-2016) was originally commissioned by the Australian Government Department of Health in 2014 (Aus $5.5 million [US $ 4.18 million]), with the broad aim of transforming the provision of mental health care across Australia by harnessing the potential of new and emerging technologies to reach all people, regardless of location, and provide them with access to timely and evidence-based treatment to improve their mental health and wellbeing [4]. Phase 1 of Project Synergy established a research and development (R&D) cycle that used participatory design methodologies to co-design, build, and evaluate a prototypic web-based platform. Project Synergy (Phase 2: 2017-2020) is another Australian Government Department of Health-funded initiative (Aus $30 million [US $ 22.81]), which is delivered by InnoWell Pty Ltd (a joint venture between the University of Sydney and PwC [Australia]) and aims to iterate the prototypic web-based prototype to the InnoWell Platform.

The InnoWell Platform links the integrated and interoperable resources (eg, apps, etools, web-based and in-clinic health services, most with data sharing functionality) to enhance service quality, track real-time health and social outcomes, and bring integrated, high-quality, and personalized service experiences to the individual seeking care. It can operate through existing health providers, such as Butterfly's National Helpline, to promote access to high-quality and cost-effective mental health services.

Importantly, the goal of the InnoWell Platform is to offer immediate web-based assessment (all individuals complete a tailored self-report questionnaire) resulting in a personalized dashboard of results. The results provide individuals with an overall profile of their health and wellbeing (including mental health), which can be shared with their health professional (HP), other health care providers, and family members, among others (dependent on permission being granted by the individual). The platform utilizes staged care based on a transdiagnostic clinical staging model [5,6] to identify the extent of disease progression at a point in time. This enables the platform to match recommendations, including apps and etools as well as clinical interventions to an individual's level of need.

At its core, the InnoWell Platform promotes person-centered health care and its principles highlight that individual clients of a service are equal partners in their health care. To that end, to promote transparency, individuals have access to all information that directly concerns them. Furthermore, all information is

presented in plain language, and individuals are presented with sufficient information to understand all components of the Platform (for example. self-report questionnaire, dashboard of results, etc), with options for obtaining further information if desired. Critically, decisions about an individual's care are made collaboratively with a HP or service, taking into account both clinical needs and personal preferences. The platform helps minimize variability in care provision between individual HPs and services by utilizing evidence and data rather than relying solely on clinical opinion, which can be variable and fallible. Finally, the platform is designed to maximize the use of resources and minimize duplication of services and wastage of time for all individuals.

A key feature of the platform is that it can be customized and configured to meet the needs of all end-users, including individuals and SO, HPs, service managers, and administrators. By engaging potential end-users through the iterative use of participatory design, the platform can be continuously developed to best meet the needs of a health care service.

Research has shown that using participatory design processes to co-design technology solutions allows for the active participation of all stakeholders and helps ensure that the end product meets the needs of its intended user base, improves usability, and increases engagement of all individuals [4,7]. Through the engagement of stakeholders in co-design, technology solutions to practical problems related to health care are generated as a means to effect reform [8]. Importantly, end-users (in this instance, all members of Butterfly's National Helpline community) have the opportunity to actively co-design the technology solution in conjunction with researchers and product designers with the aim of developing a web-based clinical tool that is more likely to be engaging and effective for all users [7,9].

The aim of the current research was to actively engage individuals from Butterfly's National Helpline community, via co-design, to collaboratively customize and configure the InnoWell Platform to enhance access to and service quality of Butterfly's National Helpline.

Methods

This research was approved by the University of Sydney Human Research Ethics Committee (Project number: 2018/041).

Participants

Participants of the participatory design workshops (detailed below in the Participatory Design Workshops section) were part of Butterfly's National Helpline community. This included individuals with a lived experience (LE) of eating disorders, disordered eating, body image and related issues, SOs, HPs (including Butterfly's National Helpline counselors), service managers, and administrators.

To be eligible to participate in the face-to-face workshops, members of the community (described above) had to be aged 15 years and older and proficient in English.

Recruitment Strategy

The recruitment strategy included the distribution of digital and nondigital postcards and A3/A4 posters with information about the participatory design workshops in the lead up to each scheduled workshop in each of the six locations. Specifically, participants were recruited via Butterfly National Helpline management, Butterfly Recovery Support Service group facilitators, targeted emails to stakeholders, and advertisements in relevant services (for example. *headspace* Darwin). To avoid any perceived coercion, recruitment was passive such that a potential participant needed to contact the Research Project Manager who, only upon a potential participant's request, then forwarded the study information sheet and participant consent form.

In line with Project Synergy's recruitment process (Reported in the study by LaMonica et al [10]), all participants were provided with detailed information about the research prior to attending a participatory design workshop. Potential participants completed a brief Screener Survey to determine eligibility for the research. Once eligibility was confirmed, potential participants were given the opportunity to provide consent. At the beginning of each workshop, the facilitators provided a second opportunity for participants to ask questions and clarify details of the research prior to providing their written informed consent. Participants were reminded that participation was entirely voluntary, and that if they agreed to participate, they could withdraw their consent at any time without being required to provide any reasons and with no impact on their relationship with The Butterfly Foundation, Butterfly's National Helpline, The University of Sydney's Brain and Mind Centre or InnoWell Pty Ltd.

Participatory Design Workshops

Participatory design workshops were run face-to-face across Australia in diverse urban and regional centers. Final locations were determined through a collaborative discussion between the Butterfly Foundation and the researchers. Locations were selected in areas where the Butterfly Foundation had a physical presence and local partnerships, and there was representation from at least one major capital city, two regional centers, and one location that was regional or otherwise isolated from eating disorder services. Six participatory design workshop locations were selected, which included two held in Sydney (major capital city), the Sunshine Coast (major city; inner regional center), Albury Wodonga (inner regional center), Darwin (outer regional), and Hobart (inner regional center). Each workshop lasted for up to four hours. As per the Project Synergy R&D Cycle [4], the series of workshops were conducted rapidly (between February and May 2019) to maintain the momentum of idea creation, continuing until theme saturation had been reached. In addition, key follow up themes based on the ideas generated in previous workshops were explored as a main focus in later workshops. This was carried out to ensure that ideas concerning these key topics had been fully saturated (Figure 1 for main focus areas). All workshops were coordinated by at least two facilitators, one of whom was a mental HP whose role was to respond to any participant concerns or distress as a result of the subject matter. A scribe was present to take detailed

hand-written notes and quotes throughout the workshop. An important component of the face-to-face participatory design workshops was that no technology was used, as research has shown that this approach results in the generation of more ideas and design solutions [11].

Figure 1. Workshop location, sample size, and focus. PD: participatory design.

Each participatory design workshop was designed to actively engage participants in interactive discussions about how to co-design potential technology solutions for Butterfly's National Helpline community and how the InnoWell Platform could be customized and configured to enhance service provision. The workshop agendas were collaboratively refined by the joint research management team, comprising researchers, HPs, and members of the Helpline community, to determine how best to discover how the technology solution might enhance or reform the service. As shown in Table 1, the workshops used a three-phased approach of discovery, evaluation, and prototyping [4]. Several specific focus areas were explored in depth in the discovery phase, which were conducted using a semistructured topic guide. Focus areas included understanding: the role of the Helpline compared to other types of services and supports (face-to-face, telemedicine, apps, and etools); how technology is used for health and wellbeing; how technology can assist individuals and SO using the Helpline; and understanding issues arising in both urban and regional areas. In the evaluation stage, a variety of methods were employed, including a review of working prototypes (wireframes) and gaining feedback on ideas generated in previous workshops through prompted discussion. In the prototyping stage methods included creation of descriptive artifacts and mock-ups, and group-based and individual development of user personas (hypothetical typical end users of the InnoWell Platform including their profile, user group, background and history, current situation, goals and motivations, frustrations, and challenges) and user journeys (series of steps illustrating how end users might interact with the Platform). As highlighted in other research [10], user journeys assist in understanding user behavior, identifying other potential areas of platform functionality for future development, defining both the taxonomy and interface, and feedback into a number of technology building activities including information architecture and sitemaps, the development of wireframes, and functional specifications.

Table 1. Participatory design workshop stages, description and focus.

Stage	Description	Methods	
		Type	Focus
Discovery	Open and prompted discussion to explore participant practices, goals, values and needs within their group and region.	Qualitative exploration using semistructured interview guides	• General views based on individual's goals, needs and values relating to the Helpline and technology • The role of the Helpline compared to other types of services and supports (face-to-face, telemedicine, apps and etools) • The use of technology for health and wellbeing • The use of technology to assist individuals and supportive others using the Helpline • Understanding issues arising in both urban and regional areas
Evaluation	Participants explore and evaluate current resources focusing on their strengths and weaknesses.	Group-based and individual review work	• Review of working prototypes (wireframes) • Feedback on ideas generated in previous workshops through prompted discussion
Prototyping	Brainstorming with participants as they suggest ideas, sketch concepts, and envision the technology solution.	Group-based and individual development work	• Descriptive artifacts and mock-ups • User personas (hypothetical typical end users of the InnoWell Platform including their profile, user group, background and history, current situation, goals and motivations, frustrations and challenges) • User journeys (series of steps illustrating how end users might interact with the Platform)

At the conclusion of the workshops, individuals with LE and SO received a web-based Aus $50 (US $ 38.01) voucher of their choice (Woolworths, Big W, Caltex, Coles, Target, Kmart, JB Hi FI, or Prepaid Mobile Recharge) as reimbursement for their time. HP and service provider participants received reimbursement only if research was conducted outside of standard working hours.

Data Analysis

As described in other research [10], at the conclusion of each workshop, the notes and quotes taken by the scribe, combined with any facilitator notes, were transcribed into a report documenting the participant background (ie, participant type such as HP or individual with lived experience) as well as the content of the discussion relative to the agenda. These reports, in combination with the visual artifacts collected during the participatory design workshops (nonidentified and presented as aggregate data to ensure confidentiality) were analyzed by the knowledge translation team, which is a group of people who represent various stakeholder backgrounds and can implement research findings into practice. The team for this research included LE representatives, mental HPs, researchers, and a co-design program manager who had experience in product management. The knowledge translation processes identified themes and key learnings to inform the customization and configuration of the InnoWell Platform for Butterfly's National Helpline. In brief, knowledge translation is an interactive process of synthesizing, exchanging, and applying knowledge [12]. With the ultimate goal being that the research findings are translated into clinical practice, organizational management, technology development, and policy reform [12].

In addition, all workshops notes, artifacts, and reports were anonymized and reviewed by three researchers (AM, AH, AR) to develop a coding framework outlining all key concepts. Data were coded in NVivo 12 software using this framework. Interpretation of the data followed established thematic techniques [13], which involved an iterative process of reading, coding, exploring the pattern and content of coded data, reflection, and discussion. Similarities and differences in opinion were examined, and differences dealt with through discussion to reach consensus. This qualitative analysis strategy has been utilized in the past literature [14-16].

When presented in the Results section, the analyzed data sources are categorized into three key areas: (1) *Notes*, which are the field notes taken by the scribe during the workshops; (2) *Prototype*, which comprises the visual artifacts developed by participants; and (3) *Report*, which is the participatory design report collated immediately after the workshop by the facilitators and the scribe that summarized the workshop findings. Further, the participant source presented in the results section in the parentheses after a quote included: (1) participants with LE of eating disorders, disordered eating, body image, and related issues; (2) participants who were HPs and identified as having a LE; (3) HPs who were clinicians but may also perform an administrative or service manager role (HP); (4) SO, or participant background not specified (PBNS). The analyzed data are presented as results in inverted commas. Data derived from the notes and artifacts are direct quotes from participants, whereas the data presented in the report may be a summary of the findings paraphrasing the participants.

Results

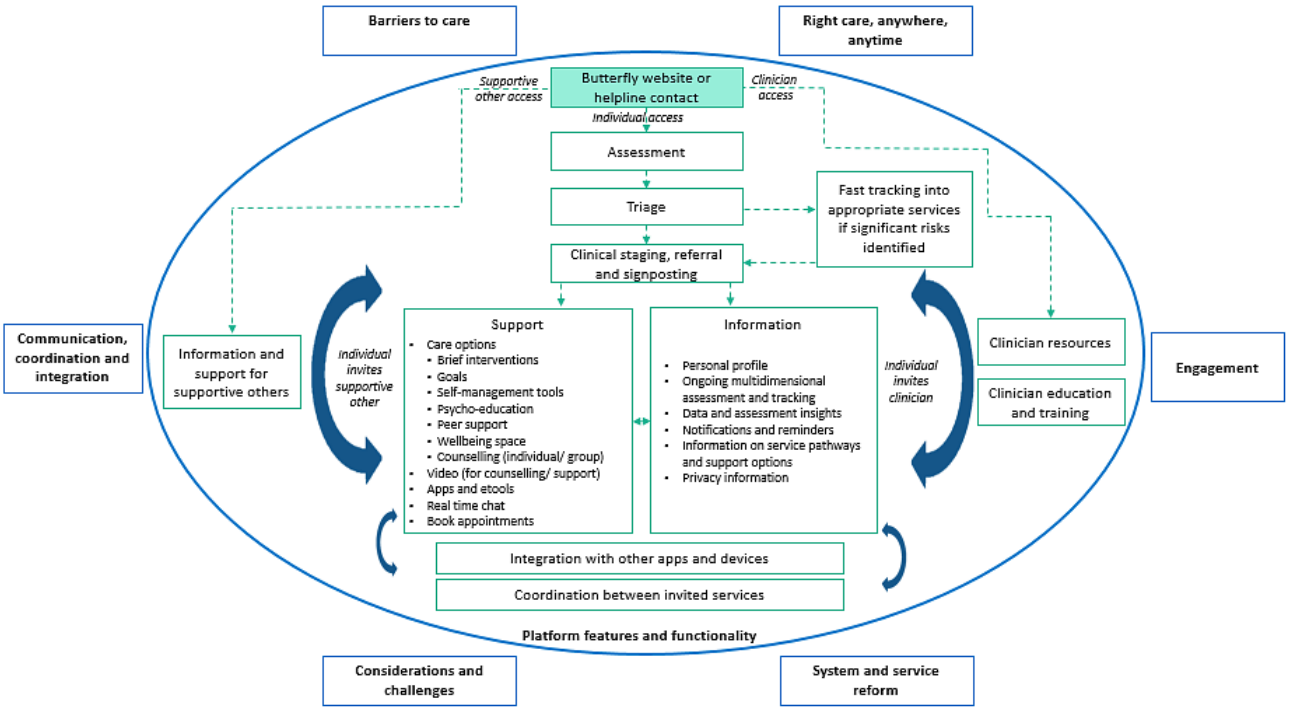
Demographics

Six participatory design workshops were held in diverse urban and regional centers across Australia, including Sydney (Workshop 1: n=9), the Sunshine Coast (Workshop 2: n=7), Albury Wodonga (Workshop 3: n=11), Darwin (Workshop 4:

n=5), Hobart (Workshop 5: n=5), and Sydney (Workshop 6: n=8). In total, 45 participants attended the workshops. Participants included people from Butterfly's National Helpline community, 13 of whom (29%) were identified as having a LE of eating disorders, disordered eating, body image, and related issues; 4 (9%) were HPs with lived experience; 21 (47%) were HPs; and 7 (16%) were SOs. Figure 1 shows the location, main focus, and participation rate across each workshop. No participants expressed concern or experienced any distress in any of the workshops.

The results presented in Figure 2 highlight the seven overarching themes identified by participants during participatory design workshops. Within the oval, the co-designed overview of the InnoWell Platform is represented, whereas outside the oval, general themes from the top to the bottom of the figure reflect before, during, and after using the co-designed InnoWell Platform. Specifically, the themes presented above the oval highlight participants' ideas concerning accessing care, which included the identified barriers to care within the current system (Theme 1: barriers to care) and the need for people to be able to access the right care, anywhere, anytime (Theme 2: Right care, anywhere, anytime). The theme within the oval comprises recommendations for the actual technological solution that was co-designed by participants, ultimately representing the InnoWell Platform features and functionality (Theme 3: features and functionality). The outside central themes relate to general recommendations if an individual uses the InnoWell Platform. These themes relay the need for communication, coordination, and integration of a technological solution when embedded in Butterfly's National Helpline (Theme 4: Communication, coordination, and integration) and the need to consider engagement and tone within the technological solution (Theme 5: Engagement). Finally, the lowermost themes relate to the identified challenges and areas to consider when implementing a technological solution in the Helpline (Theme 6: Considerations and challenges) and the potential outcomes of the technological solution embedded in the Helpline relating to system and service reform (Theme 7: System and service reform). Each theme is discussed in detail in the following results.

Figure 2. Overarching participatory design themes.



Barriers to Care

Difficulties accessing vital psychological or medical care were reported in all participatory design workshops across all participant groups (76 references). However, there were some notable differences for regional populations, which described factors associated with regionality as the most frequently reported *barrier to care*. Access for those living in regional areas that border other regional towns can also be complicated by political and health district boundaries, "...no trained clinicians in Albury, but cannot see the clinicians in Wodonga. Costs more to treat an ED [Eating Disorder] in NSW" (Notes, SO). One solution is the use of telemedicine; however, it was noted that the financial burden of care reduces the feasibility of this solution. For example, one SO highlighted that there is "...not much availability, just as expensive as seeing someone in person" (Darwin, Notes). A *lack of quality services* was described "One person claims to have expertise, other than that there's nobody up here to talk to." (Notes, SO). Unique to living in a regional area, the issue of *anonymity* was raised, "Anonymity is very important in a country town due to the stigma of help-seeking" (Report, SO). The consequences of the illness such as the beliefs of being "not sick enough" and "not deserving help" were also highlighted as *barriers to care* (Workshop 1, Sydney, Notes, HP).

Right Care, Anywhere, Anytime

The theme *right care, anywhere, anytime* was referenced in all six workshops (83 references). This theme highlighted that technology could "reduce wait times" (Prototyping, HP) by "speed[ing] up getting access to the right support" (Notes, Participant with LE) where there is "no delay in response or engagement of support...[as individuals have] often waited too long to make contact" (Prototype, PBNS). Ultimately, the InnoWell Platform could provide "...broader ways of access—virtual, Skype, irobot for excess calls so the phone is answered [by the Helpline]" (Prototype, PBNS). One HP felt that the Platform could function as "a stepping stone into the service—[where the] Platform is the first point of contact" (Prototyping, HP), which might be accessed via The Butterfly Foundation's website (Prototype, PBNS).

Ultimately, the technology could be used to "meet people where they are at" (Notes, HP) and, as illustrated in Figure 3, be able to do this at a time and place that suits their needs. For example, it was emphasized that "Being able to connect (with someone) outside of regular session times (would be important)" (Notes, SO). Additionally, the technology was viewed as needing to be free, easy to use, *respond in real time* and facilitate "Finding the expertise—wherever they are in Australia" (Notes, SO), which may be particularly beneficial for regional and rural individuals.

Figure 3. Artifact showing an individual's data connected all in one place (Workshop 2, Sunshine Coast, Prototype, Lived Experience).

Features and Functionality

The flow of *features and functionalities* that could support an end user through an imagined web-based platform, while complementing Butterfly's National Helpline were discussed the most frequently (306 references) and are presented in detail within the oval displayed in Figure 2.

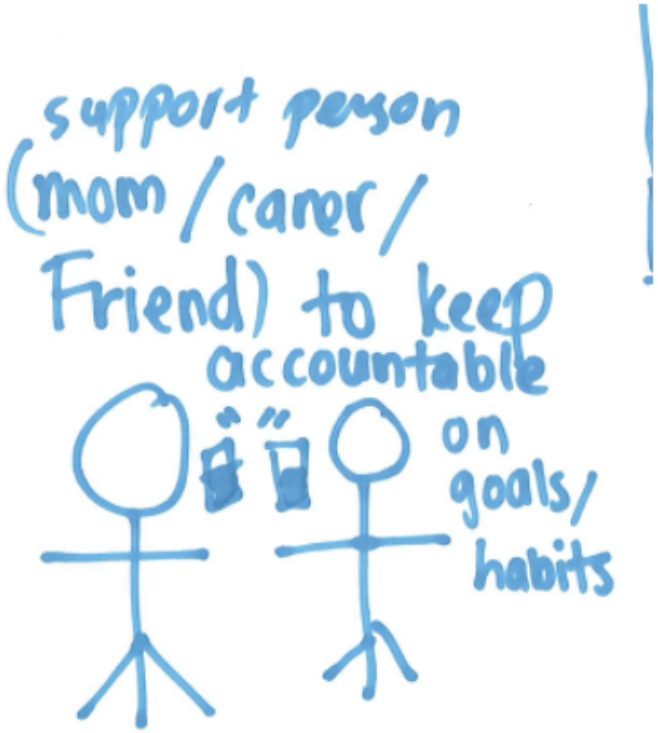
As a first step, end-users entering the InnoWell Platform would have the option of having their needs assessed. Assessment results would enable the Platform to triage the end user. If the end user presented with significant risks—such as extreme eating disorder concerns or suicidal thoughts and behaviors—additional support and fast tracking into appropriate services would be provided via the technology. If the end user did not require this, the platform would use the individual's clinical stage to give personalized referral suggestions, signpost to support, and provide the end user with relevant information. Ultimately, this was envisaged as "an automated, staging, triage platform" (Notes, HP) which "Provides support and resources based on different stages/levels of severity and distress" (Report, PBNS).

Support features included providing brief interventions (ie, care options) such as "motivational interviewing sessions" or "guided CBT (cognitive behavioral therapy)" (Notes, HP), self-management tools, peer support, goal setting, and wellbeing space. For example, one SO in Darwin suggested that the InnoWell Platform could create a space where the individual could "list the positives or their goals (eg, passions for life and hobbies). The things that will motivate them to recover" (Notes). These care options would be supported through the use of apps and etools, video counseling as "People respond so much better on the video, phone is not enough" (Notes, HP), real time chat "with someone who knows your case and can respond in real time" (Notes, SO), and appointment booking as "Sometimes it's hard for clients to make an appointment" (Notes, HP). The information features and functionality would include a personal profile of the end user, psychoeducation, information on service pathways, an easy-to-read privacy statement, and data and

assessment insights presented on the platform's dashboard of results, which provides the opportunity to complete ongoing multidimensional assessment.

The InnoWell Platform was also seen as a resource and clinical tool for HPs and SOs. Via the Platform, HPs would have access to education and training for eating disorders and body image concerns alongside resources that could be tailored by region. SOs would also have access to tailored information (such as understanding the service pathways and available resources) and support (such as peer-support and strategies for self-care). HPs and SOs could access the platform in two ways. The first is via Butterfly's National Helpline to access the above-mentioned resources. The second access pathway was via invitation from people with LE who were using the Platform themselves. This would allow HPs and SOs to provide support and be "on the same page" (Report, PBNS) as the individual they were supporting. One HP highlighted that "Bringing in a support person is key to being able to help the person" (Notes, HP). Figure 4 provides an example of how one participant with LE prototyped what this support via data sharing might look like. Importantly, it was emphasized that end users on the Platform were in control of linking and sharing their data with HPs and SOs.

Figure 4. Artifact showing data sharing with a supportive other (Workshop 2, Sunshine Coast, Prototype, Lived Experience).

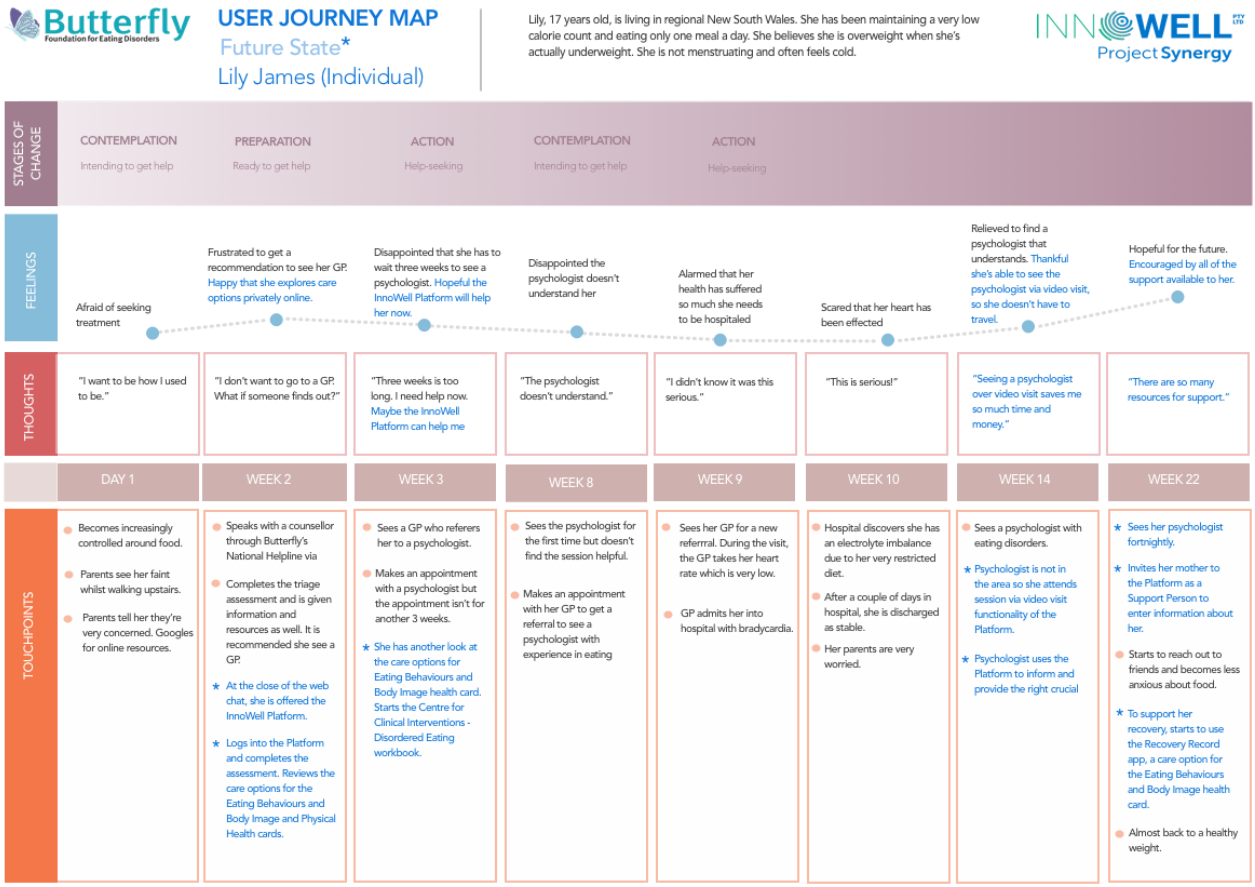


Another feature included end users being able to integrate the InnoWell Platform with other apps, devices, or medical records, and the services they were in contact with. This meant that all data and all nominated people and services involved in support of the end user were on the same page. This connection with multiple services was illustrated in an envisaged user journey, where:

> *...a young female teenager who is in the midst of a transition period (changing schools). She had a background of abuse. She contacts Butterfly's National Helpline, which connects her with the InnoWell Platform, where she has access to online resources. Through the Platform, she is connected with services. She completed a further assessment. With her health professional, she establishes goals and agrees on care options.* [Report]

Additional illustrative user journeys detailing how the imagined Platform might be used by end users were developed further by the knowledge translation team. These user journeys were based on combined participatory design workshop artifacts developed by participants in the prototyping stage of the workshops. Figure 5 presents an imagined journey of an individual with LE of eating disorders (a support person's journey is presented in Multimedia Appendix 1).

Figure 5. User journey of an individual with lived experience of an eating disorder.



Communication, Co-Ordination and Integration

Theme *communication, coordination, and integration* were frequently referenced in all six workshops (109 references). This theme highlighted the importance of the InnoWell Platform being able to assist with a collaborative care approach, where decisions were made by the individual with support from their "team" (Prototype, PBNS). This would involve *care coordination* via a multidisciplinary team approach that included allied health, "psychologists, dieticians, GPs" (Notes, PBNS), other services involved in care, and the SO where possible (as detailed above in *Platform Features and Functionality*).

A concern with the current helpline support structure was that it was delivered as a one-off session of counseling. One SO highlighted that "You need the same person, you need trust, and you need ongoing connection and ongoing care. You just need to use the existing technology. A single session is not helpful." (Notes). Ultimately, maintaining *continuity of care* for an individual was viewed as paramount. Technology was seen as potentially helping this process, as it provided a place for case conceptualization and a communication pathway. For example, a process was suggested were technology could facilitate "a case review and follow up of the caller" (Notes, PBNS) after contact with the Helpline. Another HP highlighted that "tech allows services to talk to each other" (Notes, HP), whilst an individual with LE after being shown the InnoWell Platform's dashboard of results for evaluative purposes in a later workshop stated it "Puts all [the information] in one spot which is good" and an individual using the Platform "Can show all different parties" (Notes, PBNS).

Engagement

The user experience of technology-based psychological tools and resources is largely shaped by *engagement* (150 references). Technology offers many solutions to mental health and wellbeing; however, the *need for human touch* was discussed as a priority at all workshops. The "importance of maintaining language so still a human touch, not moving too much towards bot" (Notes, HP) was raised, as were the qualities that are unique to humans: "I don't know how it replaces a hug and compassion. I do not know how tech can do that" (Notes, SO).

The impact of *language* was widely recognized as crucial to the engagement with, and tone of, the InnoWell Platform. In one workshop, it was noted "Keep language simple, broad and recovery focused. For example, include statistics of people who seek help and recovery success rates, create hope" (Report, LE). Recovery-focused language entails motivational and hope messages. Engagement would be increased by *personalized* care. Personalization includes providing choice to the users of the Helpline, "Empower consumers with choice. So much variety of what is available enables them to discern the information and choose what they want, when they want" (Report, PBNS). Difficulties in navigating credible and trustworthy information were described at all of the workshops. The "Importance of trustworthy information and resources, as well as information not being too overwhelming" (Report, PBNS) was described, otherwise "Information can be confusing

if too much or not linked to trusted resources" (Notes, PBNS). Further to information provision, trust was described as crucial to care "You need the same person, you need trust, you need ongoing connection and ongoing care" (Notes, SO).

Considerations and Challenges

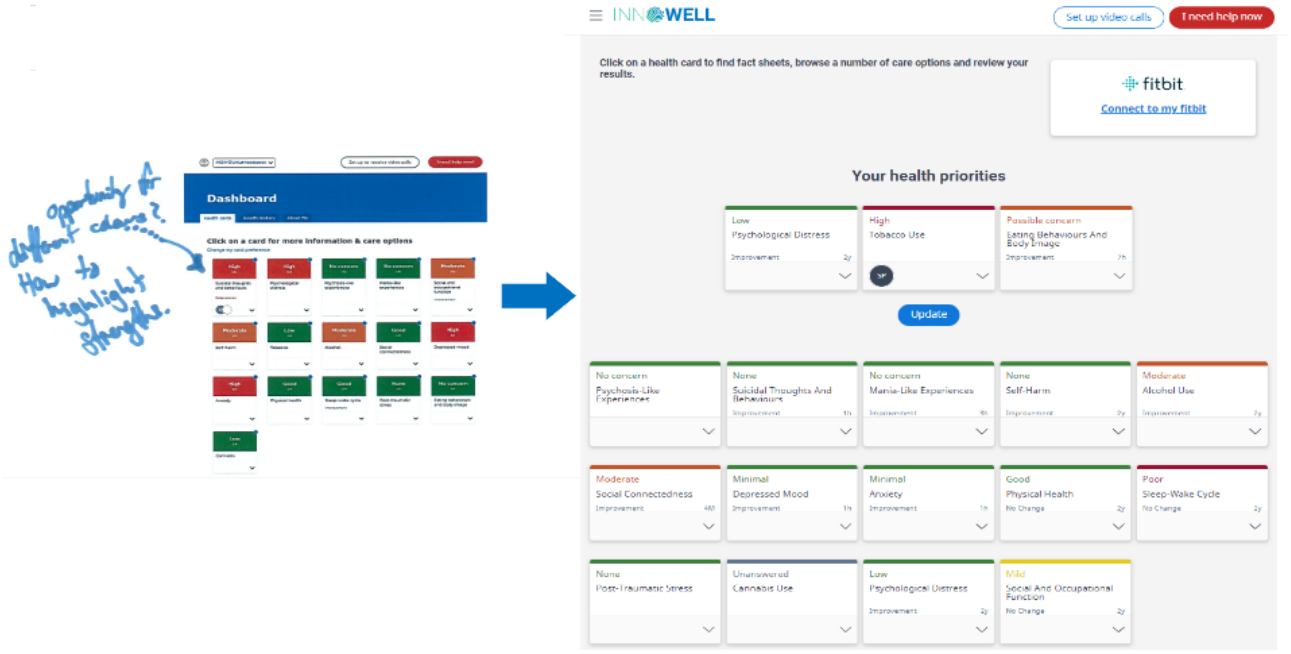
Considerations and challenges in using a technology solution (ie, The InnoWell Platform) for the Helpline were discussed in all six workshops (109 references), particularly in the Hobart and Darwin workshops.

The most frequently referenced consideration is related to the set-up of the *InnoWell Platform and infrastructure.* Chiefly in Hobart, Darwin, and the second Sydney workshop, there were *considerations and challenges* raised about *assessment, tracking, and data insights*. In terms of *assessment*, a SO highlighted that a full assessment was "too much work" (Notes, SO) and "a screener rather than a full assessment would be helpful, as it would screen the person for any major issues yet not be as time consuming" (Report, SO). Other key suggestions for the assessment were that a Butterfly counselor could "offer to start the assessment together" (Report, PBNS) with the individuals and provide breaks or chunk the assessment.

It was highlighted that in the eating disorder space, *tracking* though ongoing assessment could be problematic as "Scores could engage perfectionism" and "progress [is] hard to measure" (Notes, HP). One individual with LE highlighted that an "app actually fueled [my] eating disorder more, because [of] keeping track of things, and people with eating disorders have a desire to control their surroundings" (Notes, LE). Potential solutions to the tracking issue was that the InnoWell Platform could "Provide an option to 'opt-in' for tracking during informed consent", as well as "Communicate tracking progress/graph using the recovery model so the individual does not lose hope" and "Frame the Platform as a 'pal' (more friendly), [as it] lets you know how you are doing, and asks check-in questions" (Report, PBNS).

The third consideration relates to the need to present data insights (on a dashboard of results) in a nonconfronting and nonstigmatizing manner. For example, one SO stated that "The dashboard with a 'wall of red' would be problematic for someone who already thinks they are a piece of shit. It just confirms you have got so many problems." (Notes, SO). A HP in the same workshop suggested that data insights should present small amounts of information at one time "...to prevent visual overload" and that "...resilience and strengths need to be included to help the person feel like they have some positive aspects" In line with the *engagement* theme discussed above, an individual with LE in a later workshop highlighted that "Language used like 'Psychological Distress' [is] too confronting/strong for youth and may reinforce stigma" (Notes, LE). How feedback has been used to make changes to the InnoWell Platform to use more strength-based language and color schemes is presented in Figure 6.

Figure 6. Illustrative example of how user feedback during participatory design workshops have informed changes to the InnoWell Platform.



When an individual using the InnoWell Platform was presented with possible care options, there were multiple *resource and information considerations* highlighted by workshop participants. These related to ensuring the Platform offered existing, credible, fact-based information, which was tailored to the individuals by need and geography, and this information was actually helpful. For example, one individual with LE highlighted that "The worst thing to see is 'contact GP'. This is usually the first option but not helpful" (Notes, LE). There was a need to gather "...feedback on what works/what didn't work" in terms of the resources provided to the individual. Further, it was important for people to feel that they were not just redirected to resources, but rather supported in the options provided. A SO highlighted "it is impersonal to direct someone to another place" (Notes, SO).

Three main *service mechanisms* are highlighted for consideration. This included ensuring that there were

mechanisms to manage risk well by "...above all, do[ing] no harm" (Notes, HP), mechanisms to deal with feedback, and the need for the InnoWell Platform to operate within defined boundaries of the service. One HP in a manger position emphasized that the technology should "...focus on continually maintaining [the service] KPIs first" (Prototype, HP).

Anonymity, privacy, consent, and transparency were also important considerations. The option for people to maintain their anonymity was viewed as one of the benefits of using Butterfly's National Helpline. This benefit was also seen to carry over to the InnoWell Platform. For example, one HP highlighted that "...being online takes away the shame and is anonymous" (Notes, HP). As highlighted in the *barriers to care*, this was viewed as particularly important in rural and some regional settings, with one SO illustrating this through their statement "What applies in urban Darwin does not apply in rural communities. There is no privacy. So if you go walk into a service, everyone will know your entire mental health history." (Notes, SO).

Another key consideration related to the need for the InnoWell Platform to have the capacity to maintain confidentiality. One HP stated: "It's about one person holding the body of information. A database that holds confidentiality and can be stored securely." (Notes, HP). Furthermore, the consent process must be transparent. People with LE emphasized that "Individuals should be aware of what data is shared or escalated to the service. Ensure consent for this is provided" (Report, LE).

Considerations relating to an *individual's needs* were highlighted in four of the six workshops. This theme emphasized the need for the InnoWell Platform to be able to support individuals with complex presenting issues, those who might be resistant to support via technology as well as those of particular socio-demographics such as age and culture.

Further, it was emphasized in two workshops that the InnoWell Platform had to consider the individual's needs if a hypothetical situation arose where multiple viewpoints were included on the Platform, such as an individual with LE and their carer, but these views did not match. For example, a HP highlighted that the "...dissonance in the assessment results (eg, between carer and individual)" (Notes, HP) could be used as an opportunity for discussion. However, the individual's preferences needed to be put first, so as to address the issue raised by another HP in the same workshop, who stated that often "...carers want all [the] info, but not all individuals with LE may be wanting [a] carer perspective."

System and Service Reform

Potential future outcomes concerning the use of technology solutions, such as the InnoWell Platform, within eating disorder services generally related to the theme *system and service reform*, which was raised by participants in all workshops, irrespective of their background (132 references). This is related to the potential of technology to facilitate improvements in a *service's capacity* and the *clinical care* that could be provided.

In terms of *service capacity*, many potential outcomes are related to improvements in access to services. In one workshop, it was noted that it was "...possible that needs can be met without connecting face to face, with a clinician" (Report, PBNS). This was extended further in another workshop that imagined a scenario where technology could make wait-times more purposeful by "...re-direct[ing] them to resources whilst waiting" (Notes, HP with LE). In Darwin, a SO envisioned that technology could ensure that there was "...no delay in response or engagement support" (Notes, SO). It was envisioned that technology could make services "accessible to all" (Prototype, PBNS), particularly specific groups such as people living in regional and rural/remote areas and people who identify as "English as a second language users, males and lesbian, gay, bisexual, transgender, queer, intersex or asexual community groups" (Prototype, LE, and HP). Technology also had the capacity to improve access by "Reduce[ing] stigma and create[ing] pathways to care and support" (Report, PBNS), particularly as "Being online provides anonymity and removes stigma" (Report, PBNS).

The *clinical care* subtheme relating to *system and service reform* was referred to by participants in all workshops. This theme focused on five key areas where participants felt that technology could improve current care. The most frequently referenced area related to technology facilitates better *care coordination* (which is discussed in depth in the *communication, coordination, and integration* section). Technology could also meet care-related service gaps at follow-up as "Active efforts to follow up and transfer information is lacking" (Report, PBNS).

The next most referenced area related to improving the matching of interventions to the individual's stage of illness, which we termed *clinical staging*—this was brought up as a potential *system and service reform* outcome in five of the six workshops. In the first workshop, it was noted that technology "Can support a staged approach to care, at early stages people might just want information and to learn about skills and resources via apps and websites" (Report, PBNS). A participant with LE highlighted that it would be ideal if the technology could be used "...at any stage of the journey" (Notes, LE).

Continuity of care as a potential *system and service reform* outcome was referenced in all six workshops. For example, one HP highlighted that technology could address some key issues around the current "...lack of flexibility to find the right care and have continuity of care" (Notes, HP), with an individual with LE in the same workshop extending this line of thinking by highlighting that often people accessing services "...have to keep telling the story over and over again" (Notes, LE). Data sharing through technology allows individuals to access and show others' useful information, such as their history, current situation, support networks, and plans from one location. One HP exemplified this through their comment: "It's about one person holding the body of information" (Notes, HP). Through technology, it was envisioned that people would have a greater capacity to engage in *self-management.* It could "...help people help themselves" (Prototype, HP).

Discussion

Principal Findings

The use of participatory design processes has enabled a collaborative approach to the customization and configuration of the InnoWell Platform for Butterfly's National Helpline. This included a relatively large sample (n=45), compared to other participatory design research [16] with participants with a range of backgrounds, including individuals with lived experience, SOs, HPs, service managers and administrators, and those with mixed backgrounds. The ultimate aim of this process was to build a technology solution through prototyping that would enhance both access to and quality of care delivered through a nontraditional mental health service.

Participants felt that technology could enhance services' accessibility (ie, provide the *right care, anywhere, anytime*) and reduce the *barriers to care*, which is a finding in line with other qualitative and participatory design research in the mental health field [17,18]. Of particular importance in the current research, engagement in this participatory design process spanned both urban and regional areas. It is clear from the findings, particularly within the *barriers to care* theme, that current service provision for eating disorders is critically lacking in regional (as well as rural and remote) areas. Research has shown that in Australia, people living in regional, rural, and remote areas are likely to experience persistent disadvantages [19,20], and disadvantaged communities experience considerable social and health inequalities [20-22]. In mental health, disadvantages based on regionality are often attributed to poor access to primary and acute care services, insufficient numbers of mental health services and HPs, cost of receiving care, distance required to travel to access care, concerns about stigma, cultural barriers relating to service access, and a general reluctance to seek help in these communities [23,24]. These concerns are also highly relevant to the eating disorder community residing in regional and rural areas, as they were emphasized within our participatory design workshops as fundamental *barriers to care*.

Although technology alone cannot solve all these identified issues for regional and rural communities, participants highlighted that Butterfly's National Helpline augmenting and reforming its service using a technology solution could meet this burgeoning need to some extent. This is highlighted in the *system and service reform theme.* Further, the example user journey of a young woman with a LE of an eating disorder living in regional Australia highlights just how this might be done, using multiple forms of technology including the Helpline's webchat, the InnoWell Platform, and quality-assured apps (which may be eating disorder focused or support other areas of an individual's health and wellbeing) in conjunction with support from SOs and traditional services when they are available.

The remaining qualitative themes generated by participants in the discovery, evaluation, and prototyping stages of the workshop included communication, coordination and integration, engagement, features and functionality, and considerations and challenges. The ideas highlighted within these themes are also commonly cited in mental health focused on qualitative and participatory design research. This includes the importance of ensuring web-based tools and information are designed in such a way that they allow ease of navigation and comprehension in plain language [10,25,26]; maintain privacy and confidentiality [17,25,27,28]; ensure the information is reliable and accurate [25]; do not result in adverse impacts on the individual through use of the technology [26]; empower the individual and give them control and choice in their mental health care [10,17,25,28]; improving continuity of care within and between services [10]; support the information exchange between the clinician and the individual [10,25,26] through features and functionality such as access to assessment, progress monitoring and data tracking, data visualization, brief interventions, treatments, and reminders [10,26-28].

The InnoWell Platform is a clinical tool for clinicians (in this case, Butterfly's National Helpline counselors) for use in providing care. In the participatory design workshops, participants co-designed this clinical tool to support not only individuals accessing care via Butterfly's National Helpline, but also SO and HPs that support them (or are also supported by the Helpline). How this technology solution, which participants co-created in the prototyping stage, sits within the context of the Helpline is shown in Figure 2 and described in depth under *features and functionality*. The solution aims to promote person-centered care, which is collaborative, personalized, coordinated, and maintains continuity; to provide early intervention and support based on an individual's clinical stage; to promote self‐management through the provision of information and support-related care options; and to provide key insights based on ongoing use of the Platform's assessment. These features highlighted by participants in this study critically align with our other research on youth mental health [4,29,30], veteran populations [10], alongside other e-health solutions [31], which ultimately demonstrates the generalizability of some of these core components across both traditional and nontraditional services across the mental health sector, including eating disorder services. Ultimately, the co-designed solution adds value by demonstrating how technology can enhance a helpline service for multiple end-users with differing backgrounds. This is done by working collaboratively with end users to fully understand their needs, goals, and current practices, the context of the current support system, the challenges that may be faced when using the InnoWell Platform and how these might be addressed.

Future Directions

In terms of potential service reform outcomes, the technology solution proposed by all end users had the potential to help services meet some of the key performance indicators for public mental healthcare. As outlined by Lauriks and colleagues [32], these include ensuring clinical safety, accessibility and equity, effectiveness and outcomes, acceptability and satisfaction, efficiency, expenditure and cost, appropriateness, continuity and coordination, and workforce competence and capability [4]. These indicators will form part of our current evaluation of the implementation of the co-designed InnoWell Platform (approved by the University of Sydney Human Research Ethics Committee, protocol number: 2018/962 [33]), which aims to

determine how this technology solution is used to enhance Butterfly's National Helpline.

Limitations

A limitation of the research was that the workshops were not audio recorded. This was intentional, as dynamic individual and small group work took place making recording difficult, and to purposefully increase participant comfort and privacy. A trained scribe was presented to take detailed hand-written notes and quotes throughout the workshop, and facilitators also took notes that were triangulated for the workshop report; however, it is possible that some quotes are short-hand rather than verbatim. To value participant input, participants who did not participate during their working hours with the Butterfly's National Helpline were also offered an Aus $50 (US $ 38.01) voucher for participating. Although this amount is higher than often provided in research, the Butterfly Foundation and the research team chose this amount as it was viewed as commensurate to the 3-hour time frame of the workshops and associated travel costs (particularly in regional areas).

As a final limitation, the findings presented in this research are only a prototype. Further research is needed to understand the acceptability and usability of the platform in the context of the Helpline. An impact evaluation is currently being undertaken to assess the real-world validity of the co-designed solution.

Conclusions

The meaningful engagement of all end users from Butterfly's National Helpline community via participatory design processes demonstrated that the principles of the InnoWell Platform align with the recommendations of the National Agenda, which call for the provision of web-based prevention, early access pathways, self-help, and recovery assistance [1]. This technology solution stemming from these participatory design workshops is not an end point. The knowledge translated information gathered through stakeholder engagement across urban and regional Australia will guide the ongoing development of the platform. Ultimately, impact evaluation will provide ongoing feedback to ensure that this is a high-quality, cost-effective, evidenced-based, person-centered service for everyone affected by eating disorders or body image issues across Australia.

Acknowledgments

We would like to thank the participants of the Butterfly National Helpline 1800 ED HOPE community for their involvement in this study. We also thank the Project Synergy LE Advisory Function (LEAF) lead by Ms Sue Muller for their contribution to this research. In particular, we would like to thank LEAF members Dr Katherine Gill and Ms Leilani Darwin for co-facilitation of Workshops 3 and 4, respectively. We also thank Dr Katherine Gill for her contributions to the knowledge translation team.

Authors' Contributions

IH and TD obtained funding to support the study. The study was designed by TD and MD and conducted by MD, AH, and ACM. ACM, AH, and AER conducted the literature review, contributed to knowledge translation, and prepared the first draft of the manuscript. All authors contributed to and approved the final manuscript. The authors wish to acknowledge Lisa Whittle and Cristina Ricci for their contributions to this study.

Conflicts of Interest

IH was an inaugural commissioner on Australia's National Mental Health Commission (2012-2018). He is the Co-Director, Health and Policy at the Brain and Mind Centre (BMC) University of Sydney. The BMC operates early intervention youth services at Camperdown under contract to headspace. IH has previously supported community-based and pharmaceutical industry-supported (Wyeth, Eli Lily, Servier, Pfizer, AstraZeneca) projects focused on the identification and better management of anxiety and depression. He was a member of the Medical Advisory Panel for Medibank Private until October 2017, a Board Member of Psychosis Australia Trust and a member of the Veterans Mental Health Clinical Reference group. He is the Chief Scientific Advisor to, and an equity shareholder in Innowell. Innowell has been formed by the University of Sydney and PwC to deliver the Aus $30 (US $22.81) M Australian Government-funded 'Project Synergy'. Project Synergy is a three-year program for the transformation of mental health services through the use of innovative technologies. The remaining authors do not have any conflicts of interest to declare.

Multimedia Appendix 1
User journey of a support person.
[PNG File , 99 KB - jmir_v23i2e19532_app1.png]

References

1. The National Agenda for Disorders 2017 to 2022. The Butterfly Foundation. 2018. URL: https://thebutterflyfoundation.org.au/assets/Uploads/National-Agenda-for-Eating-Disorders-2018.pdf [accessed 2020-12-25]
2. Hayes P. Millions for research into eating disorders.: RACGP; 2019. URL: https://www1.racgp.org.au/newsgp/clinical/millions-for-research-into-eating-disorders [accessed 2020-12-24]

3. The Butterfly Foundation. About Butterfly Foundation. 2020. URL: https://thebutterflyfoundation.org.au/about-us/ [accessed 2020-12-24]
4. Hickie IB, Davenport TA, Burns JM, Milton AC, Ospina-Pinillos L, Whittle L, et al. Project Synergy: co-designing technology-enabled solutions for Australian mental health services reform. Med J Aust 2019 Oct 06;211 Suppl 7(S7):S3-S39. [doi: 10.5694/mja2.50349] [Medline: 31587276]
5. Cross S, Hickie I. Transdiagnostic stepped care in mental health. Public Health Res Pract 2017 Apr 27;27(2):1-5 [FREE Full text] [doi: 10.17061/phrp2721712] [Medline: 28474049]
6. Hickie IB, Scott J, McGorry PD. Clinical staging for mental disorders: a new development in diagnostic practice in mental health. Med J Aust 2013 May 20;198(9):461-462. [doi: 10.5694/mja13.10431] [Medline: 23682872]
7. Bjerkan J, Hedlund M, Hellesø R. Patients' contribution to the development of a web-based plan for integrated care - a participatory design study. Inform Health Soc Care 2015 Mar;40(2):167-184. [doi: 10.3109/17538157.2014.907803] [Medline: 24786524]
8. Hagen P, Collin P, Metcalf A, Nicholas M, Rahilly K, Swainston N. Participatory Design of evidence-based online youth mental health promotion, intervention and treatment.: Young and Well Cooperative Research Centre; 2012. URL: http://pandora.nla.gov.au/pan/141862/20160405-1343/www.youngandwellcrc.org.au/wp-content/uploads/2014/03/Young_and_Well_CRC_IM_PD_Guide.pdf [accessed 2020-12-24]
9. Clemensen J, Larsen SB, Kyng M, Kirkevold M. Participatory design in health sciences: Using cooperative experimental methods in developing health services and computer technology. Qual Health Res 2007 Jan;17(1):122-130. [doi: 10.1177/1049732306293664] [Medline: 17170250]
10. LaMonica HM, Davenport TA, Burns J, Cross S, Hodson S, Veitch J, et al. Technology-Enabled Mental Health Service Reform for Open Arms - Veterans and Families Counselling: Participatory Design Study. JMIR Form Res 2019 Sep 19;3(3):e13662 [FREE Full text] [doi: 10.2196/13662] [Medline: 31538937]
11. Heintz M, Law E, Soleimani S. Paper or Pixel? Comparing Paper- and Tool-Based Participatory Design Approaches. In: Human-Computer Interaction – INTERACT 2015. Switzerland: Springer, Cham; 2015 Presented at: IFIP Conference on Human-Computer Interaction; 2015; Bamberg, Germany p. 978-973. [doi: 10.1007/978-3-319-22698-9_34]
12. Pablos-Mendez A, Shademani R. Knowledge translation in global health. J Contin Educ Health Prof 2006;26(1):81-86. [doi: 10.1002/chp.54] [Medline: 16557508]
13. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
14. Milton AC, Mullan B, Hunt C. Information giving challenges and support strategies at the time of a mental health diagnosis: qualitative views from Australian health professionals. Soc Psychiatry Psychiatr Epidemiol 2016 May;51(5):735-746. [doi: 10.1007/s00127-016-1187-6] [Medline: 26880009]
15. Milton AC, Mullan BA. A qualitative exploration of service users' information needs and preferences when receiving a serious mental health diagnosis. Community Ment Health J 2015 May;51(4):459-466. [doi: 10.1007/s10597-014-9761-4] [Medline: 25027015]
16. Ospina-Pinillos L, Davenport TA, Ricci CS, Milton AC, Scott EM, Hickie IB. Developing a Mental Health eClinic to Improve Access to and Quality of Mental Health Care for Young People: Using Participatory Design as Research Methodologies. J Med Internet Res 2018 May 28;20(5):e188 [FREE Full text] [doi: 10.2196/jmir.9716] [Medline: 29807878]
17. Bucci S, Morris R, Berry K, Berry N, Haddock G, Barrowclough C, et al. Early Psychosis Service User Views on Digital Technology: Qualitative Analysis. JMIR Ment Health 2018 Oct 31;5(4):e10091 [FREE Full text] [doi: 10.2196/10091] [Medline: 30381280]
18. Ospina-Pinillos L, Davenport TA, Navarro-Mancilla AA, Cheng VWS, Cardozo Alarcón AC, Rangel AM, et al. Involving End Users in Adapting a Spanish Version of a Web-Based Mental Health Clinic for Young People in Colombia: Exploratory Study Using Participatory Design Methodologies. JMIR Ment Health 2020 Feb 08;7(2):e15914 [FREE Full text] [doi: 10.2196/15914] [Medline: 32027313]
19. McLachlan R, Gilfillan G, Gordon J. Deep and persistent disadvantage in Australia.: Productivity Commission Canberra; 2013. URL: https://www.pc.gov.au/research/supporting/deep-persistent-disadvantage/deep-persistent-disadvantage.pdf [accessed 2020-12-24]
20. Vinson T, Rawsthorne M, Beavis A, Ericson M. Dropping off the edge 2015: Persistent communal disadvantage in Australia.: Jesuit Social Services / Catholic Social Services Australia; 2015. URL: http://rcfv.archive.royalcommission.vic.gov.au/CustomData/Exhibits/TV/WIT.0103.001.0015.pdf [accessed 2020-12-24]
21. Australian Bureau of Statistics. National Survey of Mental Health and Wellbeing: Summary of Results.: Australian Bureau of Statistics; 2007. URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4326.0Main+Features32007?OpenDocument [accessed 2020-12-24]
22. Hunter E. Disadvantage and discontent: A review of issues relevant to the mental health of rural and remote Indigenous Australians. Aust J Rural Health 2007 Apr;15(2):88-93. [doi: 10.1111/j.1440-1584.2007.00869.x] [Medline: 17441816]
23. Bishop L, Ransom A, Laverty M, Gale L. Mental Health in Rural and Remote Communities.: Royal Flying Doctor Service of Australia; 2017. URL: tinyurl.com/ypbrurvs [accessed 2020-12-24]

24. National Rural Health Alliance. Mental health in rural and remote Australia. FACT SHEET - DECEMBER 2017.: National Rural Health Alliance; 2017. URL: https://www.ruralhealth.org.au/sites/default/files/publications/nrha-mental-health-factsheet-dec-2017.pdf [accessed 2020-12-24]
25. Aref-Adib G, O'Hanlon P, Fullarton K, Morant N, Sommerlad A, Johnson S, et al. A qualitative study of online mental health information seeking behaviour by those with psychosis. BMC Psychiatry 2016 Jul 11;16:232 [FREE Full text] [doi: 10.1186/s12888-016-0952-0] [Medline: 27400874]
26. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, et al. Youth Codesign of a Mobile Phone App to Facilitate Self-Monitoring and Management of Mood Symptoms in Young People With Major Depression, Suicidal Ideation, and Self-Harm. JMIR Ment Health 2018 Jan 23;5(1):e9 [FREE Full text] [doi: 10.2196/mental.9041] [Medline: 29362208]
27. Juarascio AS, Goldstein SP, Manasse SM, Forman EM, Butryn ML. Perceptions of the feasibility and acceptability of a smartphone application for the treatment of binge eating disorders: Qualitative feedback from a user population and clinicians. Int J Med Inform 2015 Oct;84(10):808-816 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.06.004] [Medline: 26113461]
28. Schmitt Z, Yarosh S. Participatory Design of Technologies to Support Recovery from Substance Use Disorders. In: Proceedings of the ACM on Human-Computer Interaction. 2018 Nov 01 Presented at: Proceedings of the ACM on Human-Computer Interaction; 2018; - p. 1-27. [doi: 10.1145/3274425]
29. Davenport TA, LaMonica HM, Whittle L, English A, Iorfino F, Cross S, et al. Validation of the InnoWell Platform: Protocol for a Clinical Trial. JMIR Res Protoc 2019 May 31;8(5):e13955 [FREE Full text] [doi: 10.2196/13955] [Medline: 31152524]
30. Rowe SC, Davenport TA, Easton MA, Jackson TA, Melsness J, Ottavio A, et al. Co-designing the InnoWell Platform to deliver the right mental health care first time to regional youth. Aust J Rural Health 2020 Apr 12;28(2):190-194. [doi: 10.1111/ajr.12617] [Medline: 32281183]
31. Monsenso. The Monsenso mHealth solution for mental illnesses. Product sheet. 2017. URL: https://www.monsenso.com/wp-content/uploads/2018/02/171220.Monsenso.Product.EN_.pdf [accessed 2020-12-24]
32. Lauriks S, Buster MC, de Wit MA, Arah OA, Klazinga NS. Performance indicators for public mental healthcare: a systematic international inventory. BMC Public Health 2012 Mar 20;12(1):214 [FREE Full text] [doi: 10.1186/1471-2458-12-214] [Medline: 22433251]
33. LaMonica H, Milton A, Braunstein K, Rowe S, Ottavio T, Jackson T, et al. Technology-Enabled Solutions for Australian Mental Health Services Reform: Impact Evaluation. JMIR Form Res 2020 Nov 19;4(11):e18759 [FREE Full text] [doi: 10.2196/18759] [Medline: 33211024]

Abbreviations

HP: health professional
LE: lived experience
PBNS: participant background not specified
R&D: research and development
SO: supportive other

Edited by G Eysenbach; submitted 23.04.20; peer-reviewed by J Bjerkan, A van Elburg; comments to author 08.06.20; revised version received 27.07.20; accepted 13.09.20; published 16.02.21.

Please cite as:
Milton AC, Hambleton A, Dowling M, Roberts AE, Davenport T, Hickie I
Technology-Enabled Reform in a Nontraditional Mental Health Service for Eating Disorders: Participatory Design Study
J Med Internet Res 2021;23(2):e19532
URL: http://www.jmir.org/2021/2/e19532/
doi:10.2196/19532
PMID:33591283

Original Paper

Awareness and Potential Impacts of the Medicalization of Internet Gaming Disorder: Cross-sectional Survey Among Adolescents in China

Yanqiu Yu[1], PhD; Ji-Bin Li[2], PhD; Joseph T F Lau[1], PhD

[1]Center for Health Behaviours Research, Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[2]Department of Clinical Research, Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Corresponding Author:
Joseph T F Lau, PhD
Center for Health Behaviours Research
Jockey Club School of Public Health and Primary Care
The Chinese University of Hong Kong
Prince Whales Hospital, Sha Tin
Hong Kong
Hong Kong
Phone: 852 22528727
Email: jlau@cuhk.edu.hk

Abstract

Background: The Eleventh Revision of *International Classification of Diseases* (ICD-11) newly listed gaming disorder, including internet gaming disorder (IGD), as a disease. The level of awareness and potential positive and negative impacts of this medicalization among adolescents were unknown.

Objective: This study investigated the levels, associated factors, and potential positive and negative impacts of awareness of the medicalization of IGD among adolescents in China.

Methods: In a cross-sectional survey, 1343 middle school students in Guangzhou, China, self-administered an anonymous questionnaire in classrooms (October to December 2019). Three risk subgroups were identified: those who scored ≥5 items in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* checklist (IGD-S), those who self-perceived having IGD currently (IGD-PC), and those who self-perceived having IGD within 12 months (IGD-P12M).

Results: Of the internet gamers, 48.3% (460/952) were aware of the medicalization of IGD; they were more likely to belong to the IGD-P12M/IGD-S risk subgroups. Within the IGD-PC/IGD-P12M (but not IGD-S) risk subgroups, IGD medicalization awareness was positively associated with favorable outcomes (reduced internet gaming time in the past 12 months, seeking help from professionals if having IGD, and fewer maladaptive cognitions). After being briefed about the ICD-11 inclusion of IGD, 54.2% (516/952) and 32.8% (312/952) expressed that it would lead to the reduction of gaming time and help-seeking behaviors, respectively; however, 17.9% (170/952), 21.5% (205/952), 15.9% (151/952), and 14.5% (138/952) perceived self-doubt for being diseased, stronger pressure from family members, negative emotional responses, and labeling effect, respectively. With a few exceptions, such perceived positive or negative impacts were stronger among the IGD-S, IGD-PC, and IGD-P12M risk subgroups.

Conclusions: The exploratory study shows that the medicalization of IGD may have benefits that need maximization and potentially harmful effects that need minimization. Future studies should test the efficacies of health promotion that increases IGD medicalization awareness.

(J Med Internet Res 2021;23(2):e22393) doi:10.2196/22393

KEYWORDS

gaming disorder; ICD-11; high-risk subgroups; disease awareness; medicalization; internet gaming; awareness; impact; adolescent; young adult; China; game; disorder; ICD

Introduction

Excessive internet gaming may cause a range of psychological and behavioral problems among adolescents [1,2]. Following the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) definition announced in 2013, the World Health Organization (WHO) listed internet gaming disorder (IGD) as a subtype of gaming disorder in the *International Classification of Diseases, Eleventh Revision* (ICD-11) in 2018 [3] and formally endorsed the decision in May 2019 [4]. The classification of IGD as a disease reflects a medicalization process, which defines a health condition as a new disease that usually requires medical treatments [5]. Medicalization of diseases (eg, attention deficit hyperactivity disorder and social anxiety disorder) has been controversial [6-8]. The same is true for IGD. Supporters of the medicalization of IGD found similar neurological changes and addictive features among people with IGD and those with substance use disorders; they believed that medicalization would advance understandings of etiology, diagnosis, and treatment of IGD [9-18]. In contrast, the researchers who disagreed with this position were concerned about the absence of evidence-based treatment, overdiagnosis, and stigma toward heavily engaged internet gamers [19-21]. Medicalization's potential benefits include new opportunities for treatments, awareness for prevention, and reduction of stigma by regarding affected people as patients instead people who are weak or have character flaws [5,9,10,17]. It is important to understand whether adolescents know about the medicalization of IGD and how they respond to it.

We contend that awareness of the official ICD-11 inclusion of IGD (represented by the term *IGD medicalization awareness* in this study) may increase positive coping behaviors that may lead to prevention (ie, reduction of gaming time in the past 12 months) and treatment (ie, intention to seek help from mental health professionals if having IGD) among adolescent internet gamers, especially those at higher risk of IGD. Such contentions have not been tested, but the belief that internet addiction is an illness was positively associated with willingness to change pathological internet habits [22]. Conceptually, IGD medicalization awareness may increase perceived severity of problematic internet gaming. Both the fear appeal theory [23] and the health belief model [24] postulate that perceived severity of a health-related problem is associated with the adoption of related preventive behaviors. In this study, at-risk adolescent internet gamers included those whose DSM-5 scores objectively exceeded the cutoff point (IGD-S) and those who subjectively perceived that they were having IGD currently (IGD-PC) or going to have IGD in the next 12 months (IGD-P12M). In the cases of internet addiction, which was significantly correlated with IGD [25], only 28.2% of those who self-perceived having internet addiction intended to correct their addiction problems [26]. Furthermore, those who were at high risk of internet addiction were even less likely than others to change their internet habits [22]. Improvements are needed to improve motivation to reduce unhealthy gaming behaviors among adolescents at risk of IGD; health promotion to increase their IGD medicalization awareness is potentially useful.

IGD medicalization awareness may alter maladaptive cognitions related to internet gaming, which are known determinants of IGD [27,28]. A study comprehensively reviewed such maladaptive cognitions and proposed a 4-factor structure that was used to construct the Internet Gaming Cognition Scale [27,29]. It was modified into a 3-factor scale (ie, the Chinese version of Revised Internet Gaming Cognition Scale), which was validated among Chinese adolescents [30] and used in this study. Those with IGD medicalization awareness might restructure their maladaptive cognitions. For instance, they might perceive internet gaming as less rewarding if they knew that it was a disease. We thus contended a negative association between IGD medicalization awareness and maladaptive cognitions related to internet gaming within the aforementioned risk subgroups of internet gamers.

Despite potential benefits, the medicalization of IGD may in parallel cause unintended negative consequences [19]. It may trigger unfavorable emotional responses among internet gamers, especially those at higher risk of IGD. According to the common sense model, illness representation that includes both cognitive and emotional representations may generate emotional responses to the disease of concern in both diseased people [31] and laypeople [32]. Hence, internet gamers with IGD medicalization awareness (especially those with self-perceived IGD) may generate negative emotions related to problematic internet gaming (eg, anxiety, guilt, blame, and shame). Second, the medicalization of IGD may induce stigma and self-stigma related to internet gaming [20,33]. Globally, people with mental illnesses encounter stigma [34]. Although the ICD-11 definition specifies that heavily engaged internet gamers who have not exhibited serious problems due to internet gaming in the past 12 months are not IGD cases [3], the general public may be unable to distinguish between heavily engaged (but healthy) gamers and IGD cases [20]. Third, parental control of adolescents' internet gaming is common and often results in adolescent-parent conflicts [35]. When parents know about the medicalization of IGD, they may exert stronger pressure on adolescent internet gamers, enhancing their perceived stress [20]. Health workers hence need to alleviate potential negative consequences while pursuing the benefits of the medicalization of IGD. Perceptions of such positive and negative consequences of the medicalization of IGD have not been investigated.

Give such background, this study investigated (1) the prevalence of IGD medicalization awareness among adolescent internet gamers in mainland China; (2) adjusted associations between IGD medicalization awareness and reduction of internet gaming time (past 12 months), intention to seek help from mental health professionals if having IGD, and IGD-related maladaptive cognitions in 3 objectively and subjectively defined high-risk subgroups (IGD-S, IGD-PC, and IGD-P12M); (3) descriptions of perceived positive impacts (eg, reducing internet gaming time) and negative impacts (eg, labeling effect, emotional distress, and stronger pressure from family members) of the medicalization of IGD after participants were briefed about the inclusion of IGD into ICD-11 by the WHO; and (4) adjusted associations between the 3 types of IGD risk status and the aforementioned postbriefing perceived impacts.

Methods

Participants and Procedure

An anonymous cross-sectional survey was conducted among grade 8 (8 years of formal education) students of 4 secondary schools selected by nonrandom sampling from October to December 2019 in Guangzhou, China. Under the supervision of trained and experienced field workers, the students self-administered the questionnaire in the classroom setting without the presence of teachers. Participants were briefed that the return of the questionnaire implied informed consent. No incentives were given. The data collection procedure was described elsewhere [30]. Of the 1343 completed questionnaires (response rate of 99.1%), 1327 (98.8%) were valid. Data obtained from the 962 (72.5%) who had played internet games in the past 12 months were analyzed. The study was approved by the survey and behavioral research ethics committee of the Chinese University of Hong Kong (No. SBRE-18-430).

Measures

Background Variables

Information about sex (male or female), living arrangement with parents (whether living with both parents, either of the parents, or neither of the parents), single-parent family status, relative household income to their classmates (much higher, higher, moderate, lower, or much lower), and self-reported academic performance (below average, average, or above average) was collected.

IGD Medicalization Awareness

The item was: “Do you know that IGD has been defined as a disease by the WHO (yes/no responses)?”

Objectively and Subjectively Defined IGD Risk Status

IGD-S was objectively defined as the endorsement of 5 or more of the 9 items of the validated Chinese version of the DSM-5 checklist [36,37]; Cronbach alpha was .74 in this study.

IGD-PC was assessed subjectively: “Do you think that you currently have IGD (yes=1, no=0)?”

IGD-P12M was subjectively assessed: “Do you think that you are going to have IGD in the next 12 months (yes=1, no=0)?” Similar questions on self-perceived IGD status have been used in previous internet addiction studies [26,38].

Maladaptive Cognitions Related to Internet Gaming

The validated 15-item Chinese version of the Revised Internet Gaming Cognition Scale has an overall scale and 3 subscales (0=never to 4=always) [30]. The overall scale was used in this report (Cronbach alpha .93).

Positive Coping Behavior/Intention

The two items, answered yes=1 or no=0, were “Have you reduced internet gaming time in the past 12 months?” and “Would you seek help from mental health professionals if you have IGD?”

Postbriefing Perceived Impacts of the Medicalization of IGD

After being briefed that “The WHO approved the ICD-11 on May 25, 2019, which defined IGD as a disease. The member states of the WHO should develop their new treatment and prevention policies prior to January l, 2022,” participants rated a 6-item checklist (yes/no responses) on whether the new ICD-11 definition of IGD (medicalization) would impact them positively (ie, leading to participants’ reduction of gaming time and seeking help from others) or negatively (ie, the news would lead to self-doubt being diseased, increase in parental pressure against playing internet games, labeling effect, and emotional distress due to playing internet games). These questions were asked at the last part of the questionnaire and thus could not affect the responses to the other questions.

Statistical Analysis

Logistic regression analyses were performed to investigate the associations involving binary outcomes, adjusted for background variables. Adjusted odds ratios and 95% confidence intervals were reported. Analysis of covariance was performed to compare between-group differences in the continuous dependent variables, adjusted for background variables. Cohen *d* represented the effect sizes of the between-group differences. SPSS Statistics 21.0 (IBM Corporation) was used for data analysis; 2-tailed $P<.05$ and $.05<P<.10$ denoted statistical significance and marginal statistical significance, respectively.

Results

Descriptive Statistics

The results are presented in Table 1. About two-thirds (601/952, 63.1%) of the internet gamers were males; 14.0% (133/952) did not live with both parents; 11% (105/952) came from single-parent families; 9.6% (91/952) perceived lower/much lower household income relative to classmates; 27.3% (260/952) self-reported below-average academic performance. Of the internet gamers, 10.8% (103/952), 58.9% (561/952), and 60.5% (576/952) belonged to the IGD-S, IGD-PC, and IGD-P12M risk subgroups, respectively (see Table 1). Within such 3 subgroups, 50.5% (52/103), 62.6% (351/561), and 65.5% (377/576) self-reported that they had reduced internet gaming time in the past 12 months (69.1% [658/952] among all gamers), and 31.1% (32/103), 39.4% (221/561), and 43.6% (251/576) reported that they would seek help from mental health professionals if having IGD (44.9% [427/952] among all gamers), respectively.

Table 1. Descriptive statistics of the participants (n=952).

Characteristics	Value, n (%)
Background variables	
Sex	
Female	351 (36.9)
Male	601 (63.1)
Living arrangement with both parents	
Yes	818 (85.9)
No	133 (14.0)
Missing data	1 (0.1)
Single-parent family status	
No	844 (88.7)
Yes	105 (11.0)
Missing data	3 (0.3)
Household income relative to classmates	
Higher/much higher	282 (29.6)
Moderate	570 (59.9)
Lower/much lower	91 (9.6)
Missing data	9 (0.9)
Self-reported academic performance	
Above average	198 (20.8)
Average	492 (51.7)
Below average	260 (27.3)
Missing data	2 (0.2)
IGD[a] status (scored or perceived)	
DSM-5[b] scored IGD	
No	845 (88.8)
Yes	103 (10.8)
Missing data	4 (0.4)
Self-perceived having IGD currently	
No	385 (40.4)
Yes	561 (58.9)
Missing data	6 (0.6)
Going to have IGD in the next 12 months	
No	364 (38.2)
Yes	576 (60.5)
Missing data	12 (1.3)
Any of the above (scored or perceived IGD)	
No	259 (27.2)
Yes	687 (72.2)
Missing data	6 (0.6)
IGD medicalization awareness	
No	471 (49.5)

Characteristics	Value, n (%)
Yes	460 (48.3)
Missing data	21 (2.2)

[a]IGD: internet gaming disorder.

[b]DSM-5: *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.*

Prevalence and Factors of IGD Medicalization Awareness

Of the internet gamers, 48.3% (460/952) reported IGD medicalization awareness (see Table 1). No background variables were significantly associated with IGD medicalization awareness (see Table 2). Adjusted for all the studied background variables, the negative association between IGD-P12M status and IGD medicalization awareness (adjusted odds ratio [AOR] 0.76, 95% CI 0.58-0.99) was statistically significant; the negative association between IGD-S status and IGD medicalization awareness was of marginal statistical significance (AOR 0.65, 95% CI 0.42-1.01; *P*=.056); that between IGD-PC status and IGD medicalization awareness was statistically nonsignificant (see Table 2).

Table 2. Factors of internet gaming disorder medicalization awareness[a] (n=952).

Characteristic	IGD[b] medicalization awareness		
	n (%)	ORu[c] (95% CI)	AOR[d] (95% CI)
Background variables			
Sex			
Female	176 (50.6)	1	—
Male	284 (48.7)	0.93 (0.71-1.21)	—
Living arrangement with both parents			
Yes	404 (50.3)	1	—
No	55 (43.3)	0.75 (0.52-1.10)	—
Single-parent family status			
No	413 (50.0)	1	—
Yes	45 (44.1)	0.79 (0.52-1.19)	—
Household income relative to classmates			
Higher/much higher	145 (52.7)	1	—
Moderate	268 (47.9)	0.83 (0.62-1.10)	—
Lower/much lower	42 (47.2)	0.80 (0.50-1.29)	—
Self-reported academic performance			
Above average	101 (52.1)	1	—
average	237 (49.1)	0.89 (0.64-1.24)	—
Below average	121 (48.0)	0.85 (0.59-1.24)	—
IGD status			
DSM-5[e] scored IGD			
No	417 (50.4)	1	1
Yes	40 (40.0)	0.66 (0.43-1.00)	0.65 (0.42-1.01)
Self-perceived having IGD currently			
No	203 (53.1)	1	
Yes	255 (46.6)	0.77 (0.59-1.00)	0.80 (0.61-1.05)
Going to have IGD in the next 12 months			
No	196 (54.4)	1	1
Yes	262 (46.5)	0.73 (0.56-0.95)	0.76 (0.58-0.99)
Any of the above (scored or perceived IGD)			
No	144 (56.3)	1	1
Yes	314 (46.7)	0.68 (0.51-0.91)	0.71 (0.53-0.96)

[a]Missing data were excluded from the analyses.

[b]IGD: internet gaming disorder.

[c]ORu: univariate odds ratio.

[d]AOR: adjusted odds ratio.

[e]DSM-5: *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.*

Associations Between IGD Medicalization Awareness and IGD and Potential Outcomes Within the Three High-Risk Subgroups

In the IGD-PC and IGD-P12M subgroups, IGD medicalization awareness was significantly associated with the reduction in gaming time in the past 12 months (AOR 1.46 and AOR 1.45, respectively) and the intention to seek professional help if having IGD (AOR 1.80 and AOR 1.91, respectively). Such associations were, however, not statistically significant in the IGD-S subgroup. Among all internet gamers, IGD medicalization awareness was significantly associated with the intention to

seek help from mental health professionals if having IGD (AOR 1.90, 95% CI 1.45-2.47); the association between IGD medicalization awareness and reduction in internet gaming time in the past 12 months was of marginal statistical significance (AOR 1.32, 95% CI 0.99-1.76; *P*=.06; see Figures 1 and 2). The adjusted analysis of covariance in Table 3 showed a similar pattern. The association between IGD medicalization awareness and maladaptive cognitions was not significant in the IGD-S subgroup but was statistically significant in the IGD-P12M subgroup (Cohen *d*=0.24, *P*=.01), marginally significant in the IGD-PC subgroup (Cohen *d*=0.18; *P*=.07), and significant among all internet gamers (Cohen *d*=0.18, *P*=.02).

Figure 1. Comparing percentages of participants self-reporting reduction in gaming time between those with and without internet gaming disorder medicalization awareness. IGD: internet gaming disorder; DSM-5: DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; AOR: adjusted odds ratio. (†: .05<*P*<.10; *: *P*<.05).

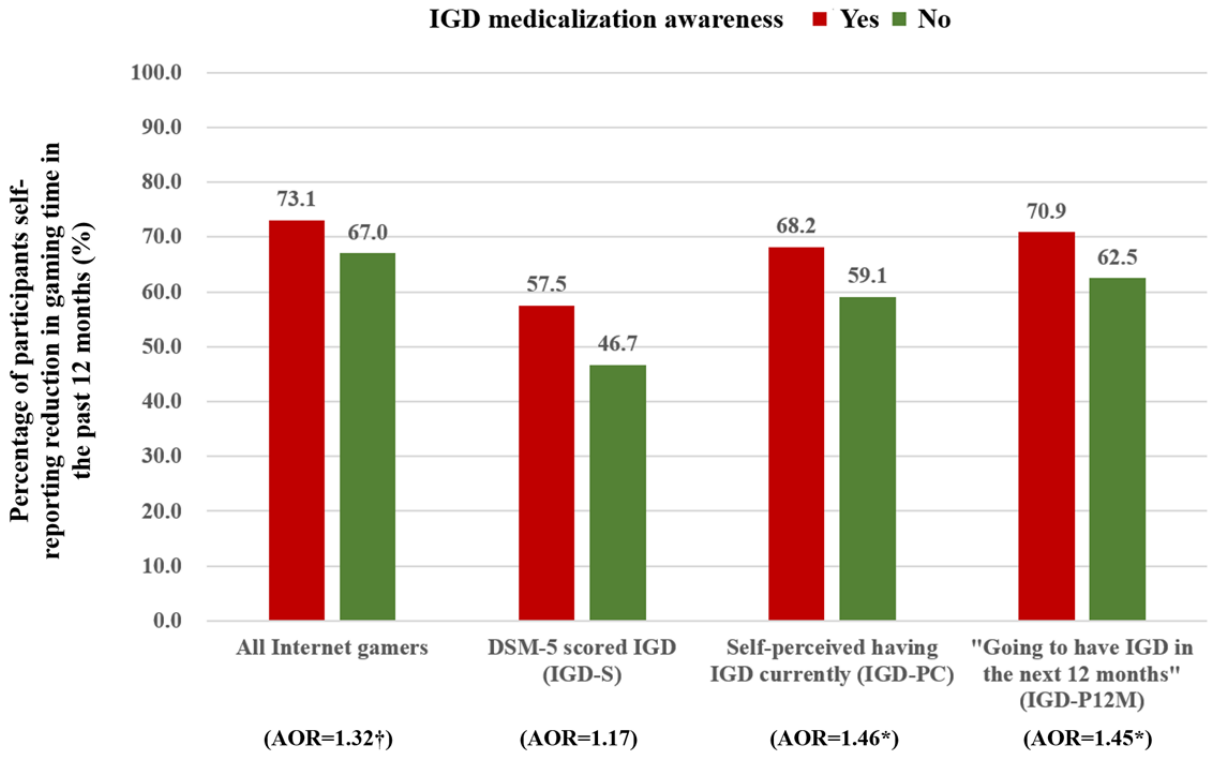


Figure 2. Comparing percentages of participants self-reporting intention to seek help from professionals between those with and without internet gaming disorder medicalization awareness. IGD: internet gaming disorder; DSM-5: DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; AOR: adjusted odds ratio. (**: *P*<.01; ***: *P*<.001).

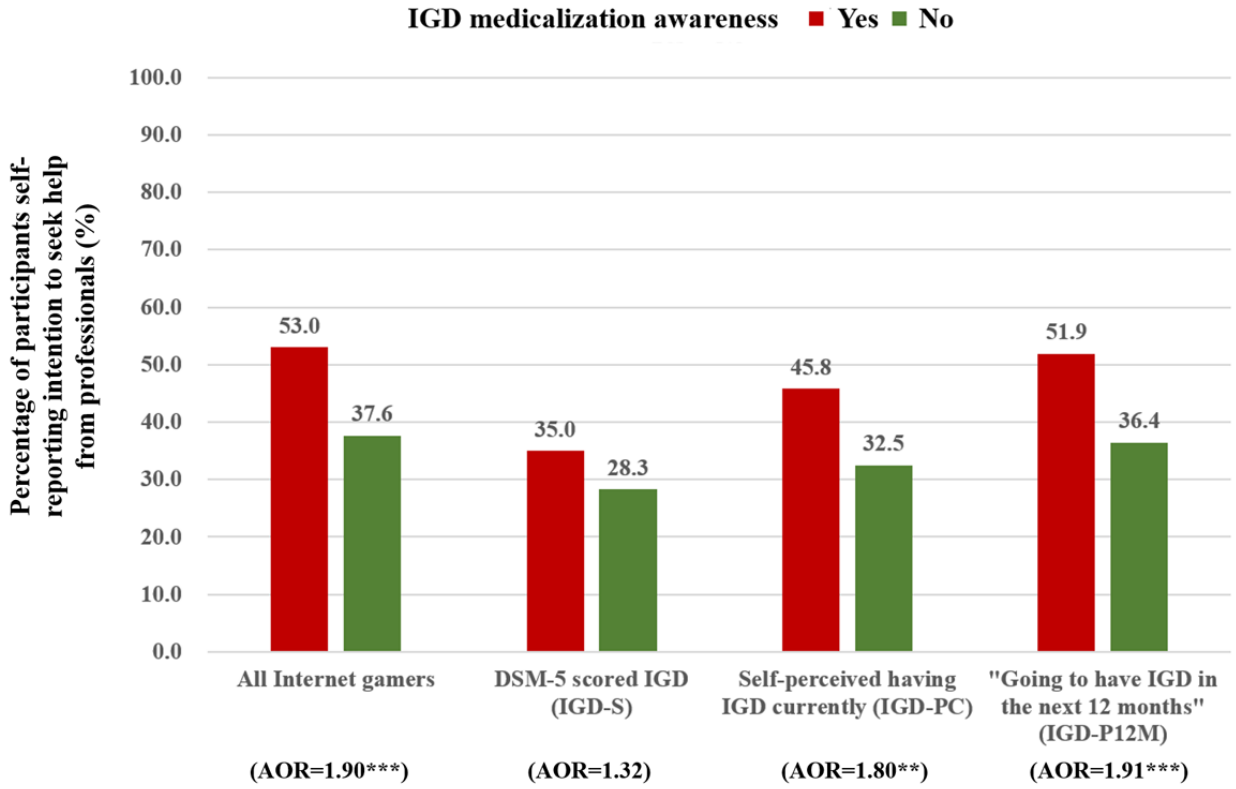


Table 3. Associations between internet gaming disorder medicalization awareness and preventive behavior/intention in the three high-risk subgroups[a].

Overall maladaptive cognitions	IGD[b] medicalization awareness			
	Yes, mean (SD)	No, mean (SD)	*P* value	Cohen *d*
DSM-5[c] scored IGD (n=100)	32.5 (14.1)	33.2 (12.0)	.82	.09
Self-perceived having IGD currently (n=582)	23.6 (11.2)	25.5 (11.4)	.07	.18
Going to have IGD in the next 12 months (n=576)	22.3 (11.0)	25.0 (11.5)	.01	.24
All internet gamers (n=952)	20.2 (11.1)	22.2 (11.8)	.02	.18

[a]Missing data were excluded from the analyses.

[b]IGD: internet gaming disorder.

[e]DSM-5: *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.*

Perceived Impacts of the Medicalization of IGD and Associations With Risk Status of IGD

After being briefed about the new inclusion of IGD into the ICD-11 by the WHO (see Measurements), 54.2% (516/952) of all the internet gamers indicated that this knowledge would make them spend less time on internet gaming, while 32.8% (312/952) indicated that it would drive them to seek help from others to deal with problems related to internet gaming (see Table 4). Besides, 17.9% (170/952), 21.5% (205/952), 15.9% (151/952), and 14.5% (138/952) of the internet gamers, after being briefed about the medicalization, perceived that it would subject them to self-doubt for being diseased, stronger pressure from family members, development of negative emotions (eg, anxiety), and label as being sick, respectively (see Table 4).

IGD-PC and IGD-P12M status but not IGD-S status were positively associated with the two perceived positive impacts (reduction in internet gaming time and intention to seek help from others to deal with problems related to internet gaming) at significant or marginally significant levels. Moreover, IGD-S, IGD-PC, and IGD-P12M status were all positively and significantly associated with the 4 types of perceived negative impacts (AOR ranged from 1.69 to 3.23) except for one association (that between IGD-P12M status and labeling effect) of marginal significance (AOR 1.50, 95% CI 1.00-2.26; *P*=.05; see Table 4).

Table 4. Perceived impacts of the International Classification of Diseases, Eleventh Revision inclusion of internet gaming disorder among internet gamers[a] (n=952).

Outcomes	IV[b]=yes, n (%)	IV=no, n (%)	AOR[c] (95% CI)
DSM-5[d] scored IGD[e]			
Intend to reduce gaming time	49 (47.6)	465 (55.0)	0.78 (0.51-1.20)
Intend to seek help from others	34 (33.0)	277 (32.8)	0.99 (0.63-1.57)
Self-doubt for being diseased	33 (32.0)	136 (16.1)	2.63 (1.62-4.26)
Stronger pressure from family members	46 (44.7)	159 (18.8)	3.23 (2.07-5.04)
Being labeled as being sick	27 (26.2)	111 (13.1)	2.23 (1.35-3.70)
Negative emotions	33 (32.0)	117 (13.8)	2.96 (1.83-4.79)
Self-perceived having IGD currently			
Intend to reduce gaming time	321 (57.2)	193 (50.1)	1.34 (1.02-1.76)
Intend to seek help from others	196 (34.9)	115 (29.9)	1.31 (0.98-1.76)
Self-doubt for being diseased	125 (22.3)	45 (11.7)	1.92 (1.31-2.83)
Stronger pressure from family members	149 (26.6)	54 (14.0)	2.01 (1.41-2.87)
Being labeled as being sick	101 (18.0)	36 (9.4)	1.94 (1.27-2.96)
Negative emotions	106 (18.9)	45 (11.7)	1.69 (1.14-2.50)
Going to have IGD in the next 12 months			
Intend to reduce gaming time	327 (56.8)	184 (50.5)	1.28 (0.98-1.68)
Intend to seek help from others	201 (34.9)	108 (29.7)	1.31 (0.98-1.75)
Self-doubt for being diseased	128 (22.2)	41 (11.3)	2.10 (1.42-3.12)
Stronger pressure from family members	157 (27.3)	45 (12.4)	2.61 (1.80-3.79)
Being labeled as being sick	95 (16.5)	40 (11.0)	1.50 (1.00-2.26)
Negative emotions	118 (20.5)	31 (8.5)	2.81 (1.82-4.33)

[a]Missing data were excluded from the analyses.

[b]IV: independent variable.

[c]AOR: adjusted odds ratio.

[d]DSM-5: *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*.

[e]IGD: internet gaming disorder.

Adjusted logistic regression models used those whose IGD risk status endorsing no as the reference groups (versus yes) and adjusted for background factors, including sex, living arrangement with both parents, single-parent family status, relative household income to their classmates, and self-reported academic performance.

Discussion

Principal Findings

In general, ICD-11 is highly influential [39]. The ICD-11 inclusion of IGD requires all nations to establish related prevention and treatment policies [9-11,13,14,16-18]. It is hence an expected driving force to reduce IGD worldwide. Health workers need to increase its benefits and reduce unintended negative consequences. It is essential to disseminate information about the new ICD-11 inclusion of IGD to adolescents and stakeholders (eg, parents, teachers, health workers, and social workers) as our data showed that the IGD medicalization awareness may reduce adolescent risky gaming behaviors and maladaptive cognitions related to internet gaming. It is equally important to understand adolescents' cognitive, behavioral, and emotional responses to the medicalization of IGD. This study filled out such knowledge gaps. There was no apparent social disparity in the IGD medicalization awareness as it was not associated with the studied background variables. Nonetheless, IGD medicalization awareness was lower in 2 high-risk subgroups (IGD-S [.05<*P*<.10] and IGD-P12M [*P*<.05]); the promotion of the disease awareness should thus target at-risk adolescents.

It is encouraging that adolescents possessing IGD medicalization awareness were more likely than their counterparts to have (1) reduced gaming time in the last 12 months, (2) intention to seek help from professionals if having IGD, and (3) fewer IGD-related maladaptive cognitions. It is plausible that the knowledge about the medicalization of IGD may have enhanced adolescents' perceived severity of playing internet games excessively and motivations to take up preventive measures (eg, reducing gaming time and seeking help) according to the fear appeal theory [23] and the health belief model [24]. Such

observed associations were triangulated by the encouraging finding that, similarly, many internet gamers indicated that they would reduce gaming time and seek help from others after being briefed about the medicalization of IGD. In the future, randomized controlled trials (RCTs) should be conducted to compare the efficacies of interventions providing adolescents IGD-related health promotion materials with and without additional information on the medicalization of IGD in fostering positive outcomes in terms of perceptions, mental distress, and behaviors related to IGD.

The associations between IGD medicalization awareness and the potential positive coping behavior/intention were more likely to be statistically significant within the 2 subjectively defined risk groups (IGD-PC and IGD-P12M) than within the objectively defined IGD group (IGD-S). The conceptual difference between diseases and illnesses is noteworthy. Diseases refer to objective clinical diagnoses, while illnesses refer to subjective experiences related to mental or physical symptoms [40,41]. The IGD-S subgroup was identified by the DSM-5 using a biomedical disease model, while the IGD-PC and IGD-P12M subgroups were subjectively evaluated and closer to the illness model. Understandably, those with illness perceptions (subjective beliefs of oneself being ill or going to be ill) were more prone to adopt positive corrective coping behaviors than those being objectively defined as IGD cases who might not feel ill. Besides, according to the health belief model [24], subjective perceptions of illness may be seen as a cue to action, which is a determinant of health-related behaviors (positive coping behaviors in our case).

Importantly, about one-fifth of the adolescent internet gamers showed concerns about side effects of the medicalization of IGD (eg, self-doubt about being diseased and worry about labeling effect). According to the common sense model [31], such problems may lead to mental health problems (eg, depression). Understandably, our data showed that the 3 at-risk subgroups were more likely than others to perceive the aforementioned negative consequences of the medicalization of IGD. To reduce such stigma, health education needs to clarify the distinction between heavily engaged internet gamers and disordered gamers.

Limitations

Although this study is possibly the first one to investigate awareness of the ICD-11 inclusion of IGD, it has some limitations. First, the findings of this study are exploratory in nature and need to be confirmed by longitudinal studies and RCTs. Second, the cross-sectional study design did not allow for the establishment of causality. Third, generalization of the results should be done with caution, as a limited number of schools were selected nonrandomly in one city in mainland China. Fourth, IGD medicalization awareness and potential responses to the medicalization (eg, reduction in gaming time) were assessed by self-reported single items that have not been validated. Fifth, social desirability bias might have inflated the levels of IGD medicalization awareness and positive coping behavior/intention. Sixth, the immediate postbriefing responses may not be reliable and may differ from actual behaviors.

Conclusions

Less than half the adolescent participants knew about the medicalization of IGD indicating there is room for improvement. The associations between IGD medicalization awareness and favorable coping behavior/intention/cognitions are encouraging. Dissemination of information about the inclusion of IGD into ICD-11 may induce adolescents to take up preventive and/or help-seeking behaviors. Such may be especially true within high-risk subgroups. Future RCTs are thus warranted to support the development of a simple, sustainable, and well-documented intervention that can be used to increase disease awareness of IGD among adolescents, possibly incorporating health promotion of healthy internet gaming. Through implementation research, such an evidence-based intervention can further be scaled up and used across countries. Furthermore, health workers need to minimize potential negative impacts of the medicalization (eg, avoidance of overpathologizing internet gamers). Research should also look at IGD medicalization awareness among other stakeholders (eg, parents, teachers, and social workers). This exploratory study is a starting point to understand the importance of potential effects of the medicalization of IGD.

Acknowledgments

This research was funded by National Natural Science Foundation of China (grant number 81373021). We would like to thank the participants for their contributions.

Authors' Contributions

YY and JTFL conceptualized the study. YY and JL designed the questionnaire. JL collected data. YY analyzed the data. YY and JTFL wrote the first draft of the manuscript. YY and JTFL revised and finalized the manuscript.

Conflicts of Interest

None declared.

References

1. Liau AK, Choo H, Li D, Gentile DA, Sim T, Khoo A. Pathological video-gaming among youth: a prospective study examining dynamic protective factors. Addiction Res Theory 2014 Dec 19;23(4):301-308. [doi: 10.3109/16066359.2014.987759]

2. Brunborg GS, Mentzoni RA, Frøyland LR. Is video gaming, or video game addiction, associated with depression, academic achievement, heavy episodic drinking, or conduct problems? J Behav Addict 2014 Mar;3(1):27-32 [FREE Full text] [doi: 10.1556/JBA.3.2014.002] [Medline: 25215212]
3. International Classification of Diseases for Mortality and Morbidity Statistics, 11th Revision. World Health Organization. 2018. URL: https://icd.who.int/browse11/l-m/en [accessed 2021-02-02]
4. World Health Assembly Update: International Statistical Classification of Diseases and Related Health Problems (ICD-11). World Health Organization. 2019 May 25. URL: http://www.who.int/news-room/detail/25-05-2019-world-health-assembly-update [accessed 2021-04-30]
5. Conrad P, Bergey M. Medicalization: sociological and anthropological perspectives. In: Wright JD, editor. International Encyclopedia of the Social & Behavioral Sciences (Second Edition). Philadelphia: Elsevier Science & Technology; 2015:105-109.
6. Nigg JT. What Causes ADHD? Understanding What Goes Wrong and Why. New York: Guilford Press; 2006.
7. Conrad P, Slodden C. The medicalization of mental disorder. In: Handbook of the Sociology of Mental Health. London: Springer; 2013:61-73.
8. Sadler JZ, Jotterand F, Lee SC, Inrig S. Can medicalization be good? Situating medicalization within bioethics. Theor Med Bioeth 2009;30(6):411-425. [doi: 10.1007/s11017-009-9122-4] [Medline: 19997778]
9. Király O, Demetrovics Z. Inclusion of gaming disorder in ICD has more advantages than disadvantages. J Behav Addict 2017 Sep 01;6(3):280-284 [FREE Full text] [doi: 10.1556/2006.6.2017.046] [Medline: 28816495]
10. Billieux J, King DL, Higuchi S, Achab S, Bowden-Jones H, Hao W, et al. Functional impairment matters in the screening and diagnosis of gaming disorder. J Behav Addict 2017 Sep 01;6(3):285-289 [FREE Full text] [doi: 10.1556/2006.6.2017.036] [Medline: 28816514]
11. Müller KW, Wölfling K. Both sides of the story: addiction is not a pastime activity. J Behav Addict 2017 Jun 01;6(2):118-120 [FREE Full text] [doi: 10.1556/2006.6.2017.038] [Medline: 28662617]
12. King DL, Gaming Industry Response Consortium. Comment on the global gaming industry's statement on ICD-11 gaming disorder: a corporate strategy to disregard harm and deflect social responsibility? Addiction 2018 Nov;113(11):2145-2146. [doi: 10.1111/add.14388] [Medline: 30007117]
13. Higuchi S, Nakayama H, Mihara S, Maezono M, Kitayuguchi T, Hashimoto T. Inclusion of gaming disorder criteria in ICD-11: a clinical perspective in favor. J Behav Addict 2017 Sep 01;6(3):293-295 [FREE Full text] [doi: 10.1556/2006.6.2017.049] [Medline: 28816497]
14. Griffiths MD, Kuss DJ, Lopez-Fernandez O, Pontes HM. Problematic gaming exists and is an example of disordered gaming. J Behav Addict 2017 Sep 01;6(3):296-301 [FREE Full text] [doi: 10.1556/2006.6.2017.037] [Medline: 28816501]
15. Rumpf H, Achab S, Billieux J, Bowden-Jones H, Carragher N, Demetrovics Z, et al. Including gaming disorder in the ICD-11: the need to do so from a clinical and public health perspective. J Behav Addict 2018 Sep 01;7(3):556-561. [doi: 10.1556/2006.7.2018.59] [Medline: 30010410]
16. Shadloo B, Farnam R, Amin-Esmaeili M, Hamzehzadeh M, Rafiemanesh H, Jobehdar M, et al. Inclusion of gaming disorder in the diagnostic classifications and promotion of public health response. J Behav Addict 2017 Sep 01;6(3):310-312 [FREE Full text] [doi: 10.1556/2006.6.2017.048] [Medline: 28816499]
17. Lee S, Choo H, Lee HK. Balancing between prejudice and fact for Gaming Disorder: does the existence of alcohol use disorder stigmatize healthy drinkers or impede scientific research? J Behav Addict 2017 Sep 01;6(3):302-305 [FREE Full text] [doi: 10.1556/2006.6.2017.047] [Medline: 28816518]
18. Saunders JB, Hao W, Long J, King DL, Mann K, Fauth-Bühler M, et al. Gaming disorder: its delineation as an important condition for diagnosis, management, and prevention. J Behav Addict 2017 Sep 01;6(3):271-279 [FREE Full text] [doi: 10.1556/2006.6.2017.039] [Medline: 28816494]
19. van Rooij A, Ferguson C, Colder Carras M, Kardefelt-Winther D, Shi J, Aarseth E, et al. A weak scientific basis for gaming disorder: let us err on the side of caution. J Behav Addict 2018 Mar 01;7(1):1-9 [FREE Full text] [doi: 10.1556/2006.7.2018.19] [Medline: 29529886]
20. Aarseth E, Bean A, Boonen H, Colder Carras M, Coulson M, Das D, et al. Scholars' open debate paper on the World Health Organization ICD-11 Gaming Disorder proposal. J Behav Addict 2017 Sep 01;6(3):267-270 [FREE Full text] [doi: 10.1556/2006.5.2016.088] [Medline: 28033714]
21. Statement on the WHO ICD-11 list and the inclusion of gaming. European Games Developer Federation. 2018 Jun 15. URL: http://www.egdf.eu/wp-content/uploads/2018/06/Industry-Statement-on-18-June-WHO-ICD-11.pdf [accessed 2021-02-02]
22. Jiang Q, Leung L. Effects of individual differences, awareness-knowledge, and acceptance of internet addiction as a health risk on willingness to change internet habits. Soc Sci Comput Rev 2011 Feb 28;30(2):170-183. [doi: 10.1177/0894439311398440]
23. Peters GY, Ruiter RA, Kok G. Threatening communication: a critical re-analysis and a revised meta-analytic test of fear appeal theory. Health Psychol Rev 2013 May;7(Suppl 1):S8-S31 [FREE Full text] [doi: 10.1080/17437199.2012.703527] [Medline: 23772231]

24. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
25. Király O, Griffiths MD, Urbán R, Farkas J, Kökönyei G, Elekes Z, et al. Problematic internet use and problematic online gaming are not the same: findings from a large nationally representative adolescent sample. Cyberpsychol Behav Soc Netw 2014 Dec;17(12):749-754. [doi: 10.1089/cyber.2014.0475] [Medline: 25415659]
26. Lau JT, Wu AM, Cheng K, Tse VW, Lau MM, Yang X. Prevalence and factors of self-corrective intention among Hong Kong secondary school students who are self-assessed Internet addiction cases. Child Adolesc Ment Health 2018 Sep;23(3):155-163. [doi: 10.1111/camh.12219] [Medline: 32677295]
27. King DL, Delfabbro PH. The cognitive psychology of Internet gaming disorder. Clin Psychol Rev 2014 Jun;34(4):298-308 [FREE Full text] [doi: 10.1016/j.cpr.2014.03.006] [Medline: 24786896]
28. Billieux J, Potenza M, Maurage P, Brevers D, Brand M. Cognitive factors associated with gaming disorder. In: Verdejo-Garcia A, editor. Cognition and Addiction. Philadelphia: Elsevier; 2020:221-230.
29. King DL, Delfabbro PH. The cognitive psychopathology of internet gaming disorder in adolescence. J Abnorm Child Psychol 2016 Feb 15;44(8):1635-1645. [doi: 10.1007/s10802-016-0135-y] [Medline: 26875565]
30. Yu Y, Mo PK, Zhang J, Li J, Lau JT. Validation of the Chinese Version of the Revised Internet Gaming Cognition Scale among Adolescents in China: Maladaptive Cognitions as Potential Determinants of Internet Gaming Disorder. Int J Environ Res Public Health 2019 Dec 31;17(1):290 [FREE Full text] [doi: 10.3390/ijerph17010290] [Medline: 31906174]
31. Chan RC, Mak WW. Common sense model of mental illness: Understanding the impact of cognitive and emotional representations of mental illness on recovery through the mediation of self-stigma. Psychiatry Res 2016 Dec 30;246:16-24. [doi: 10.1016/j.psychres.2016.09.013] [Medline: 27644016]
32. Martin R, Rothrock N, Leventhal H. Common sense models of illness: implications for symptom perception and health-related behaviors. In: Suls J, Wallston KA, editors. Blackwell Series in Health Psychology and Behavioral Medicine. Social Psychological Foundations of Health and Illness. Hoboken: Blackwell Publishing; 2003:199-225.
33. Van Rooij AJ, Schoenmakers TM, Vermulst AA, Van Den Eijnden RJ, Van De Mheen D. Online video game addiction: identification of addicted adolescent gamers. Addiction 2010 Sep 15;106(1):205-212. [doi: 10.1111/j.1360-0443.2010.03104.x] [Medline: 20840209]
34. Rüsch N, Angermeyer M, Corrigan P. Mental illness stigma: concepts, consequences, and initiatives to reduce stigma. Eur Psychiatry 2005 Dec;20(8):529-539. [doi: 10.1016/j.eurpsy.2005.04.004] [Medline: 16171984]
35. Schneider LA, King DL, Delfabbro PH. Family factors in adolescent problematic Internet gaming: a systematic review. J Behav Addict 2017 Sep 01;6(3):321-333 [FREE Full text] [doi: 10.1556/2006.6.2017.035] [Medline: 28762279]
36. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington: American Psychiatric Association; 2013.
37. Ko C, Yen J, Chen S, Wang P, Chen C, Yen C. Evaluation of the diagnostic criteria of Internet gaming disorder in the DSM-5 among young adults in Taiwan. J Psychiatr Res 2014 Jun;53:103-110. [doi: 10.1016/j.jpsychires.2014.02.008] [Medline: 24581573]
38. Pontes HM, Szabo A, Griffiths MD. The impact of Internet-based specific activities on the perceptions of Internet addiction, quality of life, and excessive usage: a cross-sectional study. Addict Behav Rep 2015 Jun;1:19-25 [FREE Full text] [doi: 10.1016/j.abrep.2015.03.002] [Medline: 29531976]
39. ICD-11 Implementation or Transition Guide. Geneva: World Health Organization; 2019. URL: https://dxrevisionwatch.files.wordpress.com/2019/06/icd-11-implementation-or-transition-guide_v105-1.pdf [accessed 2021-02-02]
40. Last J, International Epidemiological Association. A Dictionary of Epidemiology. New York: Oxford University Press; 1995.
41. Wikman A, Marklund S, Alexanderson K. Illness, disease, and sickness absence: an empirical test of differences between concepts of ill health. J Epidemiol Community Health 2005 Jun;59(6):450-454 [FREE Full text] [doi: 10.1136/jech.2004.025346] [Medline: 15911638]

Abbreviations

AOR: adjusted odds ratio
DSM-5: *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
ICD-11: *International Classification of Diseases, Eleventh Revision*
IGD: internet gaming disorder
IGD-S: Scored IGD status
IGD-P12M: Perceived future IGD status in the next 12 months
IGD-PC: Perceived current IGD status
RCT: randomized controlled trial
WHO: World Health Organization

Edited by G Eysenbach; submitted 10.07.20; peer-reviewed by A Dominguez-Rodriguez, S Stefanov; comments to author 01.10.20; revised version received 12.11.20; accepted 17.01.21; published 24.02.21.

Please cite as:
Yu Y, Li JB, Lau JTF
Awareness and Potential Impacts of the Medicalization of Internet Gaming Disorder: Cross-sectional Survey Among Adolescents in China
J Med Internet Res 2021;23(2):e22393
URL: https://www.jmir.org/2021/2/e22393
doi:10.2196/22393
PMID:33625362

Original Paper

Tracking COVID-19 Discourse on Twitter in North America: Infodemiology Study Using Topic Modeling and Aspect-Based Sentiment Analysis

Hyeju Jang[1,2], MSc, PhD; Emily Rempel[2], MSc, PhD; David Roth[2], MSc, PhD; Giuseppe Carenini[1], MSc, PhD; Naveed Zafar Janjua[2,3,4], MBBS, MSc, DrPH

[1]Department of Computer Science, University of British Columbia, Vancouver, BC, Canada

[2]British Columbia Centre for Disease Control, Vancouver, BC, Canada

[3]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

[4]Centre for Health Evaluation and Outcome Sciences, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Naveed Zafar Janjua, MBBS, MSc, DrPH
British Columbia Centre for Disease Control
655 West 12th Avenue
Vancouver, BC, V5Z 4R4
Canada
Phone: 1 604 707 2514
Email: naveed.janjua@bccdc.ca

Abstract

Background: Social media is a rich source where we can learn about people's reactions to social issues. As COVID-19 has impacted people's lives, it is essential to capture how people react to public health interventions and understand their concerns.

Objective: We aim to investigate people's reactions and concerns about COVID-19 in North America, especially in Canada.

Methods: We analyzed COVID-19–related tweets using topic modeling and aspect-based sentiment analysis (ABSA), and interpreted the results with public health experts. To generate insights on the effectiveness of specific public health interventions for COVID-19, we compared timelines of topics discussed with the timing of implementation of interventions, synergistically including information on people's sentiment about COVID-19–related aspects in our analysis. In addition, to further investigate anti-Asian racism, we compared timelines of sentiments for Asians and Canadians.

Results: Topic modeling identified 20 topics, and public health experts provided interpretations of the topics based on top-ranked words and representative tweets for each topic. The interpretation and timeline analysis showed that the discovered topics and their trend are highly related to public health promotions and interventions such as physical distancing, border restrictions, handwashing, staying home, and face coverings. After training the data using ABSA with human-in-the-loop, we obtained 545 aspect terms (eg, "vaccines," "economy," and "masks") and 60 opinion terms such as "infectious" (negative) and "professional" (positive), which were used for inference of sentiments of 20 key aspects selected by public health experts. The results showed negative sentiments related to the overall outbreak, misinformation and Asians, and positive sentiments related to physical distancing.

Conclusions: Analyses using natural language processing techniques with domain expert involvement can produce useful information for public health. This study is the first to analyze COVID-19–related tweets in Canada in comparison with tweets in the United States by using topic modeling and human-in-the-loop domain-specific ABSA. This kind of information could help public health agencies to understand public concerns as well as what public health messages are resonating in our populations who use Twitter, which can be helpful for public health agencies when designing a policy for new interventions.

(J Med Internet Res 2021;23(2):e25431) doi:10.2196/25431

KEYWORDS

COVID-19; Twitter; topic modeling; aspect-based sentiment analysis; racism; anti-Asians; Canada; North America; sentiment analysis; social media; discourse; reaction; public health

Introduction

Worldwide, more than 31 million people have been diagnosed with COVID-19, and more than 1 million people have died as of October 12, 2020 [1]. Waiting for the development and rollout of a vaccine, governments across the world have implemented wide-ranging nonpharmaceutical interventions such as hand hygiene, face masks, contact tracing, isolation and quarantine, and physical (social) distancing through banning mass gatherings and lockdowns to reduce the transmission of SARS-CoV-2. The impact of COVID-19 and measures to prevent transmission has generated a lot of discussion among the general population, medical and public health professionals, and government officials [2,3]. Some of this discourse is happening on social media such as Twitter.

During this pandemic, people have been using social media such as Twitter to share news, information, opinions, and emotions about COVID-19 [4,5], similar to previous infectious disease outbreaks such as Ebola. In the Eloba outbreak, public health organizations helped contain Ebola by monitoring conversations on social media and spreading accurate information about the disease [6-9]. As we can see from these past successes, social media is an important source to learn about people's reactions and concerns. This information can assist public health authorities in the monitoring and surveillance of health information, concerns, and behaviors, and designing interventions to reduce the impact of the pandemic. Understanding people's information needs, misinformation, hate speech and discrimination, compliance with preventative measures, and other reactions to COVID-19, and where their concerns lie helps to tailor public health strategy and ultimately create better informed interventions.

Topic modeling and sentiment analysis have been widely used to identify issues and people's opinions in public health and is being used to understand COVID-19–related issues as well (Table 1). Analyses were conducted to identify patterns of health communications in diverse kinds of data sources, communities, and locations. Although some works investigated news articles [10] or research papers [11], most research focused on social media such as Reddit posts [12] and tweets [13-19]. Conversations in particular communities were examined, such as tweets posted by US governors and presidential cabinet members [13], and African American twitter communities [16]. Specific languages and locations were discussed as well (eg, Chinese news articles [10], Persian and Farsi tweets in Iran [14], and English tweets in California and New York in the United States [17]). Although all these works investigated people's reactions toward COVID-19, there have been few studies about general public responses in Canada. Furthermore, although sentiment analysis has been broadly applied [15-18], the techniques used in prior work determine the sentiment of an overall text rather than capturing opinions toward COVID-19–specific aspects chosen by domain experts and exploit lexicon built in general domains, overlooking that a word's sentiment depends on the domain or context where it is used [20].

Table 1. Related work on topic modeling and sentiment analysis on COVID-19–related data.

Authors	Source	Posters	Time	Location	Language	Sentiment
Liu et al [10]	News articles	News reporters	January 1 to February 20, 2020	Not specified	Chinese	No
Dong et al [11]	Research papers	Researchers	Unknown to March 20, 2020	Not specified	English	No
Stokes et al [12]	Reddit posts	Public	March 3-31, 2020	Not specified	English	No
Sha et al [13]	Tweets	State governors, presidential cabinet members, and the president	January 1 to April 7, 2020	US	English	No
Hosseini et al [14]	Tweets	Public	March 13 to April 19, 2020	Iran	Persian and Farsi	No
Sharma et al [15]	Tweets	Public	March 1-30, 2020	Not specified	English	Yes
Odlum et al [16]	Tweets	Public (African Americans)	January 21 to May 3, 2020	Not specified	English	Yes
Wang et al [17]	Tweets	Public	March 5 to April 2, 2020	California and New York, US	English	Yes
Abd-Alrazaq et al [18]	Tweets	Public	February 2 to March 15, 2020	Not specified	English	Yes
Ordun et al [19]	Tweets	Public	March 24 to April 9, 2020	Not specified	English, Spanish, Italian, French, and Portuguese	No
This study	Tweets	Public	January 21 to May 31, 2020	Canada and US	English	Yes

Our study aims to investigate Twitter users' reactions to COVID-19 in North America, especially in Canada. We analyzed COVID-19–related tweets with topic modeling and aspect-based sentiment analysis (ABSA) using human-in-the-loop and interpret the results with public health experts. We examined the sentiment of tweets about COVID-19–related aspects such as social distancing and masks by using ABSA based on domain-specific aspect and opinion terms. The key advantage of our study is that public health experts are actively involved in the computational process with the specific goal of informing public health interventions. Our results were interpreted by these public health experts, and we used a human-in-the-loop ABSA approach to obtain domain specific aspect and opinion terms. To the best of our knowledge, we are the first to directly identify sentiment of COVID-19–specific aspects.

Methods

Data and Data Processing

We used a public Twitter data set about the COVID-19 pandemic, collected by Chen et al [21] using numerous COVID-19–related keywords such as "coronavirus," "COVID-19," and "pandemic." The data collection started on January 28, 2020 (tweets from January 21, 2020), and is still ongoing, which has published over 123 million tweets as of May 11, 2020. This data set includes retweets, quoted tweets, and replies to tweets.

For our study, we collected tweets until the end of May 2020, the end of the first wave in Canada, since we aim to investigate people's reactions and concerns in the early days of COVID-19. We selected tweets whose location is Canada or the United States.

Among the 372,711 tweets in total (Canada: n=30,235, US: n=342,476), we only included tweets written in English using tweet metadata and the spacy-langdetect toolkit [22]. This process resulted in 319,524 tweets in total, 25,595 for Canada, and 293,929 for the United States. To remove tweet-specific keywords and URLs, we used the tweet-preprocessor toolkit [23]. We did not remove hashtags and mentions because they can be informative for our study. We lowercased and tokenized using the Spacy toolkit [24]. Since the methods we used in this paper are all unsupervised, we did not split the data for training and test. Our scripts are available on GitHub [25].

Topic Modeling

We first discovered topics in COVID-19–related tweets using a widely used topic modeling approach, latent Dirichlet allocation (LDA) [26]. We chose to use LDA because it is simple and popular. We also tried another popular topic modeling method, nonnegative matrix factorization, but LDA results were more distinct in categories according to public health experts' assessment. As we have seen the potential of topic modeling in this study, we will also consider more sophisticated topic modeling algorithms such as pachinko allocation or hierarchical LDA, which allow modeling relations between topics. To assess changes in topics of discussion over time, we compared timelines of topic distributions and timing of public health intervention implementations for COVID-19.

To discover topics and track the topic change over time, we constructed topic models on our Twitter data using LDA implementation in the scikit-learn package [27]. We chose a model with 20 topics among 5, 10, 20, and 50 because 20 topics showed diverse and less redundant topics when manually examined.

The topics generated by LDA were interpreted and labeled by two public health experts. Both experts have extensive experiences in public health with doctoral training in the field. In the initial phase of the study before choosing a final model, they discussed the results to build consensus. After the final output was obtained, the junior expert interpreted and labeled it first, and the senior expert reviewed.

To analyze the dynamics of public health relevant topics, we investigated the change in the prevalence of the topics over time. More specifically, we performed a basic analysis based on an examination of the estimates of θ, a document-to-topic distribution, produced by the model. We first divided tweets into weekly buckets using Coordinated Universal Time–12 time stamps (eg, January 21-26, January 27 to February 2, and February 3-9, 2020). We then computed a mean θ vector for tweets in each bucket as done by Griffiths and Steyvers [28].

ABSA

To capture sentiment revealed in tweets toward important aspects of COVID-19, we used ABSA. In our study, aspects can include public health interventions or issues associated with COVID-19, such as "social-distancing," "reopening," and "masks." We investigated people's opinion (positive and negative) toward these aspects.

We used ABSApp, a weakly-supervised ABSA system [29]. We chose ABSApp because it does not require labeled data for training and allows manually editing domain-specific aspect and opinion lexicons produced by the method. This feature is particularly beneficial for us because, in collaboration with domain experts, we could select and add aspects public health agencies are interested in.

The two public health experts who labeled topics from topic modeling also edited the terms so that aspect terms are related to important public health interventions or issues they are interested in and that opinion terms are words that describe sentiment of those public health terms. Similarly to the topic interpretation process, the junior expert edited the terms first, and the senior expert reviewed.

Results

Context

Figures 1 and 2 provide context for our results, presenting mobility and case counts for Canada and the United States. These data show that, as the daily COVID-19 cases increased, activities such as recreational or work-related mobility drastically decreased in the middle of March 2020. Around the middle of March, public health measures were put in place as well. The dates of public health orders differed by provinces or

states as well as the specifics of the orders, but in Canada, on March 11, 2020, health officials in British Columbia underlined the importance of social distancing and urged people to stay home as much as possible. On March 16, 2020, an order prohibiting gatherings of 50 people or more was placed. In Ontario, a state of emergency was declared on March 17, 2020, and social distancing measures commenced. On the same day, the federal government announced closure of the Canada-US border to all nonessential traffic. In the United States, mandatory stay-at-home orders were issued, beginning in California on March 19 followed by many other states afterwards.

Figure 1. Mobility and case count for Canada from February 15 to May 31, 2020. Google mobility data is only available since February 15. pharm.: pharmacy.

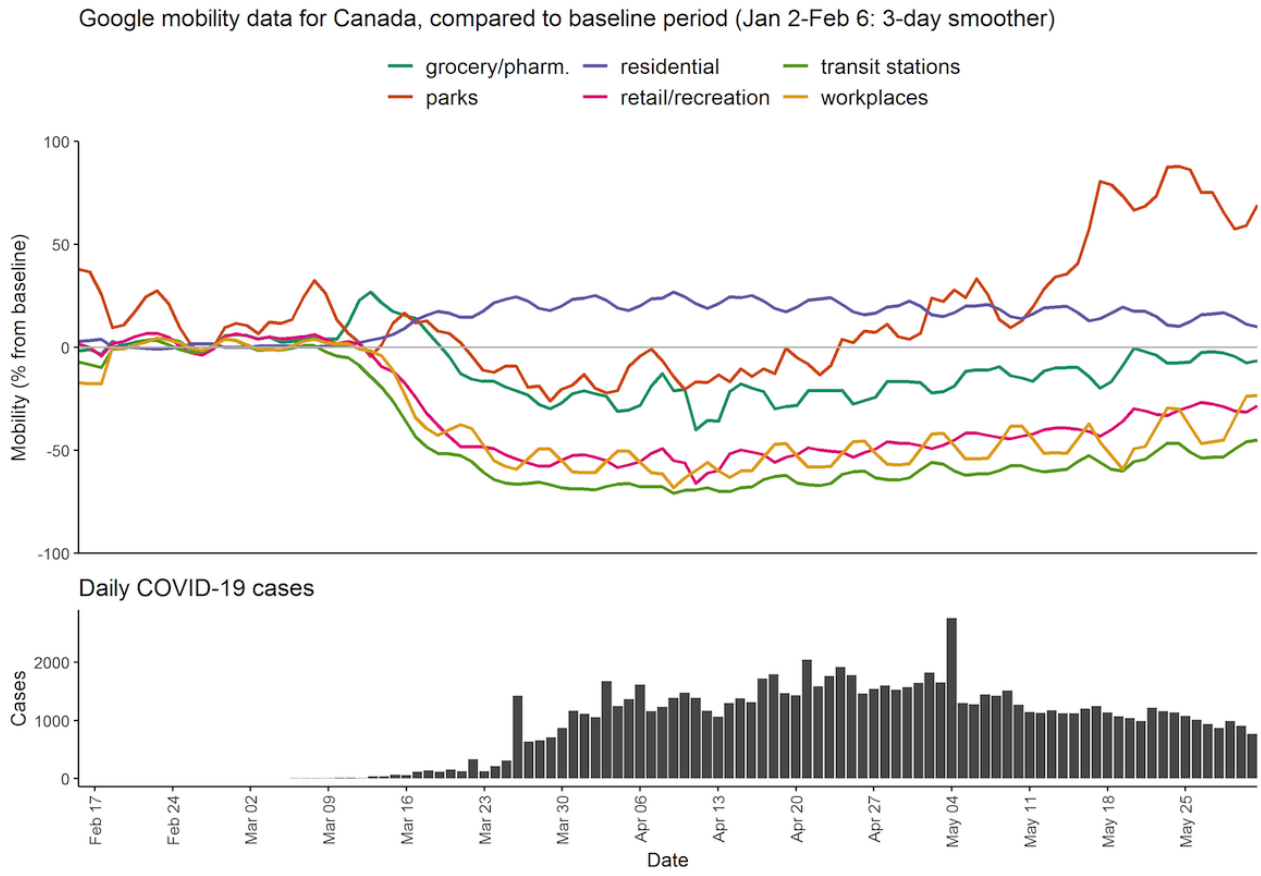
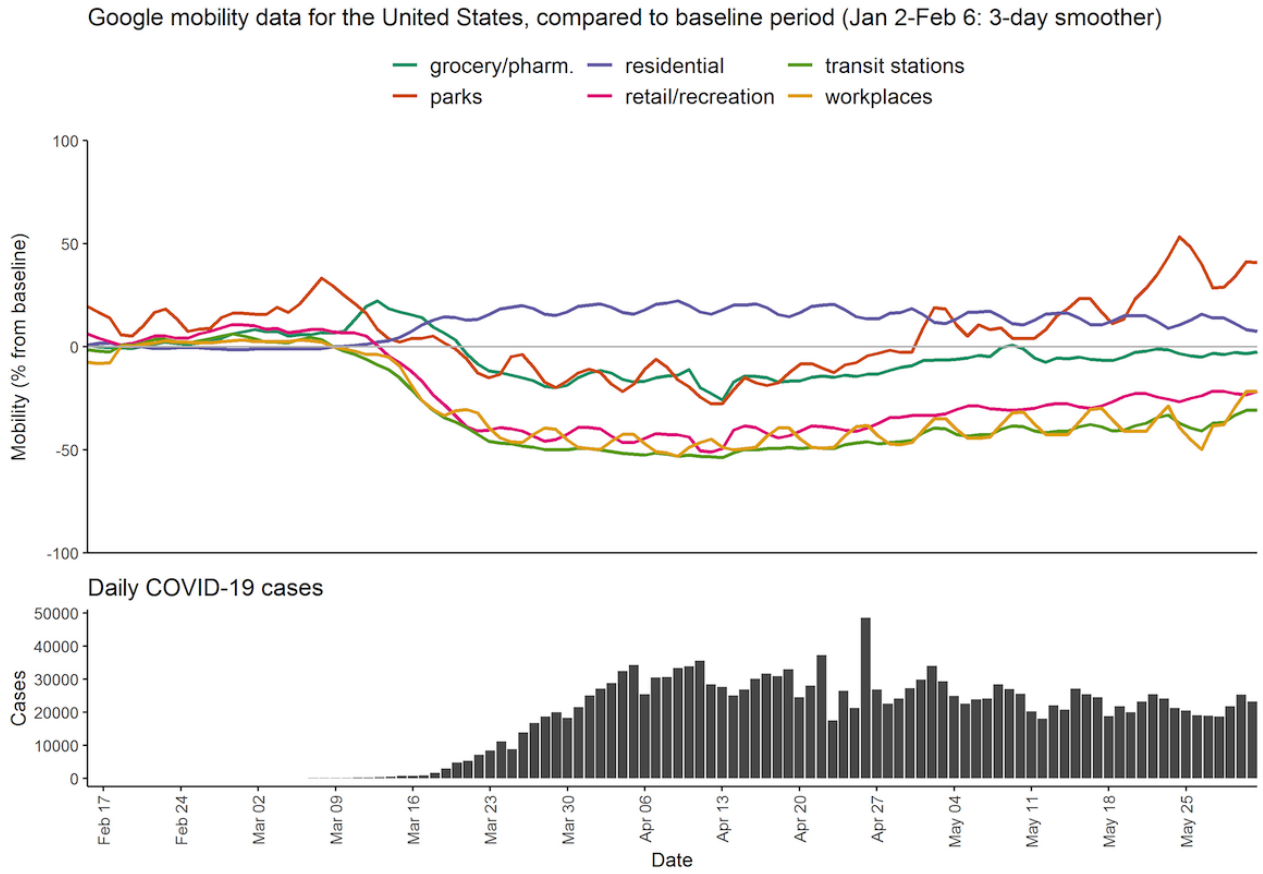


Figure 2. Mobility and case counts for the United States from February 15 to May 31, 2020. Google mobility data is only available since February 15. pharm.: pharmacy.



Topic Modeling

The discovered topics were highly related to public health promotions and interventions such as physical distancing, border restrictions, handwashing, staying home, and face coverings, as shown in Textbox 1. Other topics included US President Donald Trump, initial outbreaks in Wuhan, economic concerns, and negative reactions. The entire set of topics is listed in Multimedia Appendix 1.

Textbox 1. Top 5 prevalent topics in Canada and the United States.

Canada

1. Age and COVID-19 transmission, as well as time
2. Initial outbreak in Wuhan
3. US President Trump's statement
4. Thank you notes related to the pandemic mixed with discussion of cruise ship outbreaks
5. Air travel and regional border restrictions and outbreaks

United States

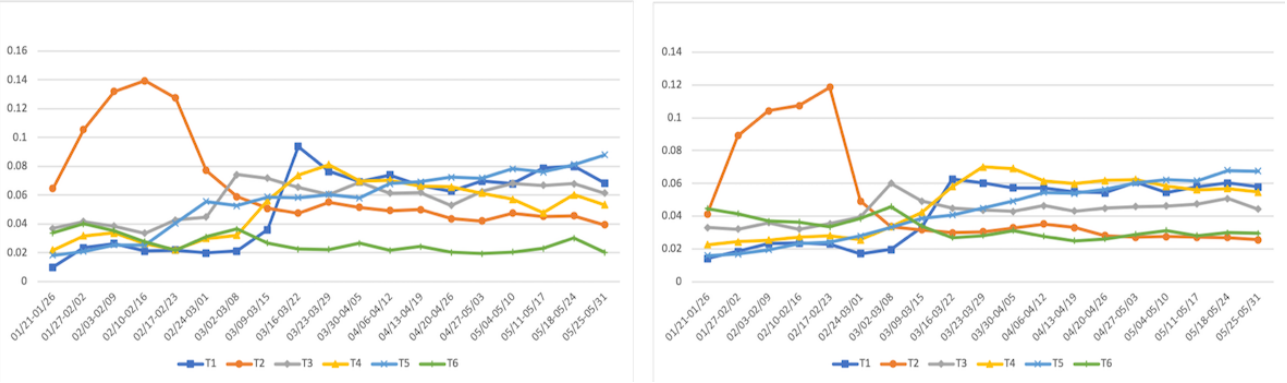
1. Age and COVID-19 transmission, as well as time
2. US President Trump's statement
3. Early debate on whether COVID-19 is like the flu.
4. Initial outbreak in Wuhan
5. The need to stay home and the impact of COVID-19 on essential workers and family

The most prevalent topics in Canada and the United States showed some differences, as can be seen in Textbox 1. In both countries, age and COVID-19 transmission was the most prevalent topic. The discussion around the initial outbreak in Wuhan and US President Trump's statement was also active in both countries. However, the topic about air travel and regional border restrictions was highly ranked only in Canada, whereas the topic was not even listed in the top 10 in the United States.

Similarly, the topics about COVID-19 being like the flu and staying home were highly ranked in the US tweets, but ranked lower than other topics in the Canadian tweets.

Based on the mean θ vector for each bucket, we drew graphs of public health–relevant topics over time as shown in Figure 3. First, we observed that the patterns in the US tweets and Canadian tweets were similar. Although there were slight differences, the overall increase and decrease patterns were almost identical. For example, the topic about air travel and regional border restrictions (T2) shows a peak in February and drastically decreases.

Figure 3. Changes of six public health–relevant topics over time. T1: social and physical distancing; T2: air travel and regional border restrictions and outbreaks; T3: handwashing and preventive measures; T4: the need to stay home and impact of COVID-19 on essential workers and family; T5: number of tests and cases; T6: masks and face coverings.



Second, we could see that the topic trend is highly related to public health interventions. For example, the topic about social distancing (T1) started to increase in early March 2020 after social distancing measures were enacted. Handwashing (T3) also started to be emphasized then. The topic about the need to stay home (T4) started to increase around the end of March. In Canada, the Federal Quarantine Order was issued on March 24, and in the United States, many states issued stay-at-home orders around that time as well. Discussion about the number of tests and cases (T5) gradually increased. Interestingly, the topic about masks and face coverings (T6) slightly decreased from March; this is possibly because public health institutes in both countries announced their position about masks around that time.

Aspect-Based Sentiment Analysis

After training the tweet data using ABSApp, we obtained 806 aspect terms and 211 opinion terms. Manually editing the lexicons resulted in 545 aspect terms (eg, “vaccines,” “economy,” and “masks”) and 60 domain-specific opinion terms such as “infectious” (negative) and “professional” (positive). These manually edited terms were then used for the inference of sentiments for 20 key aspects selected by public health experts. The results are shown in Figures 4 and 5. Overall, the sentiments between Canada and the United States showed similar patterns. We observed that the sentiments about COVID-19 itself was dominantly negative. With this, the Twitter users’ reactions to misinformation appeared to be more negative than positive, suggesting the frustration about the situation and misinformation. The mixed sentiments about masks might reflect the conflicting messaging around using masks. The negative sentiments toward Asians may imply that anti-Asian sentiments escalated due to COVID-19.

Figure 4. Aspect-based sentiment analysis results. x-axis: selected aspects; y-axis: number of positive occurrences and number of negative occurrences in log scale.

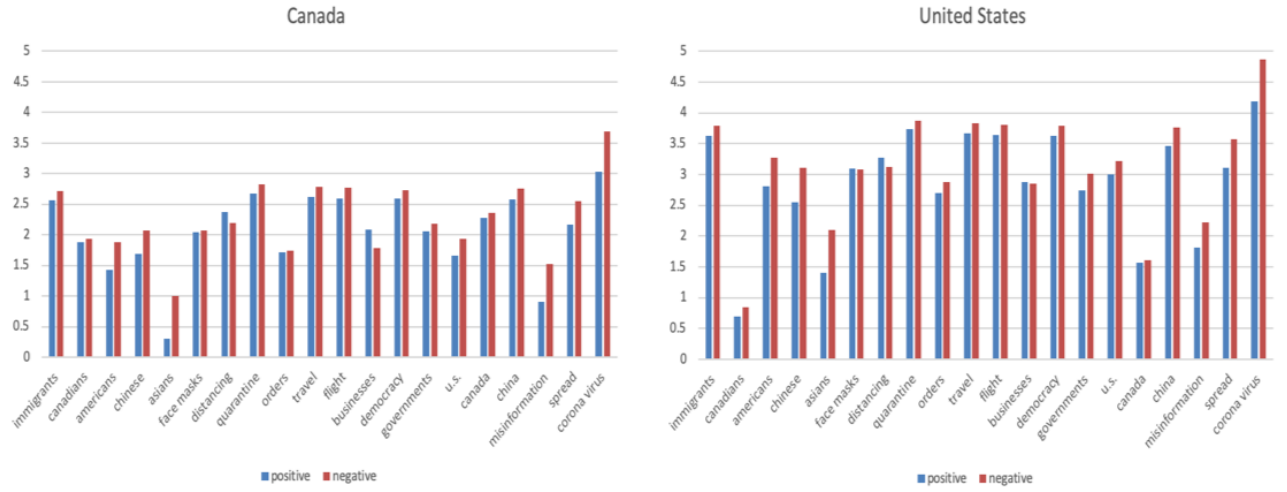
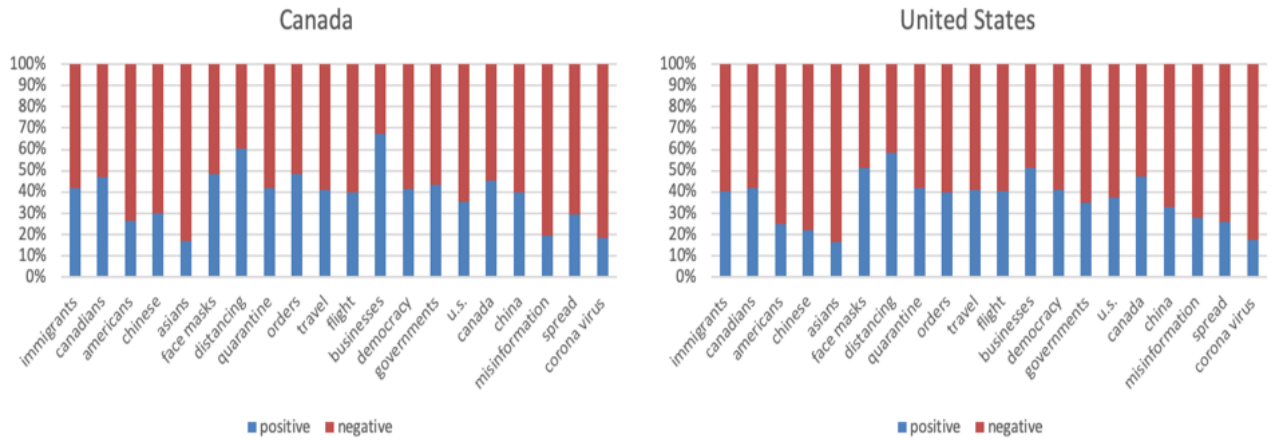


Figure 5. Aspect-based sentiment analysis results for selected aspects. y-axis: the ratio between number of positive occurrences and number of negative occurrences.



To further investigate the possible stigma for Asians, we observed words that frequently co-occurred with the aspect words Chinese and Asians. The top-ranked words in negative tweets included “virus,” “racist,” “racism,” “fucking,” “attacks,” “ass,” “assaults,” “blame,” and “hate,” and the top-ranked words in positive tweets included “fucking,” “racism,” “respectful,” “kind,” “street,” “disgusting,” and “crying.” We list sample tweets that show positive and negative sentiments in Textbox 2.

Textbox 2. Sample tweets showing positive or negative sentiments toward Asians.

Positive

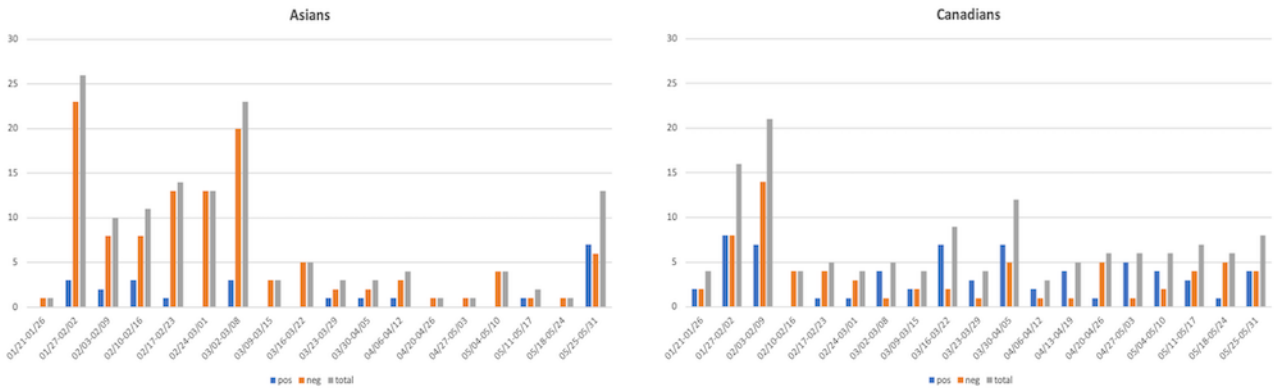
- “You should not be afraid of Asians but you should be absolutely terrified of the PEOPLE THAT DONT COVER THEIR MOUTHS/NOSES DURING A COUGH AND/OR SNEEZE.”
- “French Asians hit back at racism with 'I'm not a virus”
- “Y’all realize that the coronavirus ain’t exclusive to Chinese people right?? mfs look for any excuse to be racist bruh”

Negative

- “Oriental Asians always starting some fuckin outbreak...”
- “Yea I’m holding my breath round all Asians till this coronavirus shit clear up call it wat u think it is.”
- “No Asians allowed in my shop after the outbreak.”

Figure 6 displays sentiment changes over time toward Asians and Canadians. Although sentiments about Canadians were overall more positive except in February 2020 and after the middle of May, sentiments about Asians were mostly negative. Especially in the beginning of March when COVID-19 started to be serious in North America, we can see a spike in the number of negative tweets about Asians, and then it drastically reduces after that, which might suggest that there were some campaigns or awareness about anti-Asian racism.

Figure 6. Sentiment changes over time for Asians and Canadians. Y-axis shows the number of positive occurrences, number of negative occurrences, and number of total occurrences. neg: negative; pos: positive.



Discussion

Principal Findings

In this study, using topic modeling and ABSA on Twitter data from North America, we identified various topics related to physical distancing, travel and boarder restrictions, handwashing and preventive measures, face masks, stay-at-home orders, and the number of cases and testing. Travel and border restrictions were major discussion points in February 2020, which were taken over by other topics such as physical distancing later in time. ABSA analysis identified various negative themes related to the overall outbreak, anti-Asian racism and misinformation, and positive occurrences related to physical distancing. These data demonstrate Twitter users' focus on discussing and reacting to public health interventions during the first phase of the pandemic.

This kind of information could help public health agencies to understand public concerns as well as what public health messages are resonating in our populations who use Twitter. For example, public health agencies in North America have focused their messaging around encouraging hand hygiene, limiting physical contact when sick, and staying home to prevent infection. We can see this messaging echoing in the topics around handwashing, staying home, mask wearing, and social or physical distancing.

For public health decision makers, it would be beneficial to have the pipeline where a computational model keeps running on social media data as a stream, and the results are reviewed by public health experts. This will then be reflected in public health education communication or messages to address misinformation related to the topics.

Risk communication and knowledge translation in practice is a combination of proactive and reactive messaging [30,31]. Topic modeling can inform social media priorities and form a key rapid response function for public health communicators. Two key functions can be imagined: identifying new topics for public health communicators and assessing uptake of public health messages. The first function of developing new topics would follow the traditional knowledge translation cycle (in other words, identifying a knowledge gap, assessing barriers to use, developing products, disseminating, and iterating). When topic modeling identifies discourse either on aspects public health is not messaging on or misunderstandings of current health messages, this can start the knowledge translation cycle. In particular, this will help identify new and emerging areas for misinformation messaging. The second function is assessment of public health message uptake. We can explore whether key public health messages are showing up in social media discourse. If not, we can explore what is showing up instead.

Our findings that tweets reflect public health interventions are aligned with other studies. Abd-Alrazaq et al [18] performed topic modeling on tweets before mid-March 2020, and their results focused more on the virus itself (eg, its origin, impact on people, and the economy) but did not show conversations about public health interventions. However, in the studies using tweets from March and April, topics related to social distancing policies such as school closure, stay-at-home orders, and work from home commonly emerged in tweets posted by US governors and presidential cabinet executives [13]; Reddit posts [12]; tweets in English, Spanish, Italian, French, and Portuguese [19]; tweets in California and New York [17]; and tweets in Iran [14].

Depending on tweets used for analysis, other studies report some interesting topics different from topics drawn from tweets in Canada. For example, topics related to government and political issues were observed in the studies on tweets by US governors [13] and on tweets in Iran [14], whereas our analysis only showed Trump's statement as a topic rather than overall political issues regarding COVID-19.

Our ABSA provides sentiments toward specific aspects by considering sentence structures, while most prior works performing sentiment analysis use algorithms to decide a sentiment of an entire text. For this reason, these studies are generally not suitable for identifying a sentiment of a given aspect. For instance, Wang et al [17] computed the average sentiment scores of tweets by each day and each hour rather than obtaining sentiments for aspects. Yin et al [32] related sentiment for each tweet to the topic the tweet belonged to and then investigated the overall sentiment of each topic. Therefore, it is not straightforward to compare our ABSA results with other sentiment analysis results.

However, our ABSA results, especially related to racism and discrimination against Asians, were also observed in other research using different study methods. Zhu [33] qualitatively analyzed 1366 tweets to examine swears around "Chinese virus" in multiple languages. Topic modeling on English tweets in March and April 2020 [34] showed a topic related to racism with top-ranked words such as "Chinese" and "pig." A survey in the United States also showed prejudicial attitudes among Americans toward Chinese Americans [35]. These findings show that ABSA has the potential to track stigma and other negative consequences related to COVID-19. Our communities of Asian ethnicity have experienced unprecedented stigma and discrimination due to COVID-19. Chinese Canadians and other East Asians are experiencing hatred expressed as assaults, verbal threats, and feeling unsafe in the society. As our analysis suggests, if we monitor the change in discrimination over time using social media as a stream in real time, we could develop counteracting messages and measures in specific geographic areas whenever there is a spike in such incidents.

Our study had the following limitations. We used only a small set of Twitter data because tweets with the location information were limited compared to the whole data set. This has affected other studies using social media data in a similar fashion. Moreover, it should be noted that the geo-tagged tweets data set comprises statements from a nonuniform subsample of the population. According to Gore et al [36], only 15% of online adults regularly use Twitter, and those aged 18-29 years and minorities tend to be more highly represented on Twitter than in the general population.

In our data set, we looked at location at the country level (ie, Canada or the United States). However, Gore et al [36] showed that there could be significant geographic bias at the city level

in the sentiment expressed in tweets over the same time period. Therefore, there may be a risk that specific geographic areas at the city level might be overrepresented for a given country in our study.

Another possible bias comes from not knowing who tweeted from the locations. Padilla et al [37] showed that the sentiment of tweets could be biased based on if people are local or visiting an area at the time of their tweets. Our data set could be biased in this regard. However, given travel restrictions and use of country instead of city, this bias may not be an issue for this analysis.

In general, whenever our proposed pipeline would be deployed in practice, all these biases should be carefully considered and addressed.

In addition, although ABSA allows capturing more nuanced sentiments toward specific aspects, it also has the limitation that current state-of-the-art sentiment analysis techniques have: it cannot properly handle figurative languages such as sarcasm. However, since our proposed approach can process substantial amounts of twitter data, it should be able to deal with the noise generated by these complex pragmatic phenomena.

Conclusion

In this paper, we present the exploratory results of topic modeling and ABSA on COVID-19–related tweets in North America, especially in Canada. We compared topic modeling and ABSA results of Canada and the United States, and showed public health intervention–related topic changes over time. Our analyses demonstrated that Twitter conversations about COVID-19 are highly aligned with public health interventions. In our study, public health experts were actively involved in the computational process as well as interpretation of the results. The human-in-the-loop ABSA allowed manually editing aspect and opinion lexicons, and as a result, our analysis showed sentients toward the aspects public health experts were interested in by leveraging the domain-specific lexicons. Our results suggest that monitoring Twitter user's reactions about COVID-19–related aspects can be beneficial for public health policy makers.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Latent Dirichlet allocation–generated topics and their interpretations.
[DOCX File , 15 KB - jmir_v23i2e25431_app1.docx]

References

1. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2020-10-12]
2. Wahbeh A, Nasralah T, Al-Ramahi M, El-Gayar O. Mining physicians' opinions on social media to obtain insights into COVID-19: mixed methods analysis. JMIR Public Health Surveill 2020 Jun 18;6(2):e19276 [FREE Full text] [doi: 10.2196/19276] [Medline: 32421686]
3. Rao HR, Vemprala N, Akello P, Valecha R. Retweets of officials' alarming vs reassuring messages during the COVID-19 pandemic: implications for crisis management. Int J Inf Manage 2020 Dec;55:102187 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102187] [Medline: 32836644]
4. Yu M, Li Z, Yu Z, He J, Zhou J. Communication related health crisis on social media: a case of COVID-19 outbreak. Curr Issues Tourism 2020 Apr 14:1-7. [doi: 10.1080/13683500.2020.1752632]
5. Zhao Y, Cheng S, Yu X, Xu H. Chinese public's attention to the COVID-19 epidemic on social media: observational descriptive study. J Med Internet Res 2020 May 04;22(5):e18825 [FREE Full text] [doi: 10.2196/18825] [Medline: 32314976]
6. Carter M. How Twitter may have helped Nigeria contain Ebola. BMJ 2014 Nov 19;349:g6946. [doi: 10.1136/bmj.g6946] [Medline: 25410185]
7. Dalrymple KE, Young R, Tully M. "Facts, Not Fear": negotiating uncertainty on social media during the 2014 Ebola crisis. Sci Commun 2016 Jun 22;38(4):442-467. [doi: 10.1177/1075547016655546]
8. Guidry JP, Jin Y, Orr CA, Messner M, Meganck S. Ebola on Instagram and Twitter: how health organizations address the health crisis in their social media engagement. Public Relations Rev 2017 Sep;43(3):477-486. [doi: 10.1016/j.pubrev.2017.04.009]
9. Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? Am J Infect Control 2015 Jun;43(6):563-571. [doi: 10.1016/j.ajic.2015.02.023] [Medline: 26042846]
10. Liu Q, Zheng Z, Zheng J, Chen Q, Liu G, Chen S, et al. Health communication through news media during the early stage of the COVID-19 outbreak in China: digital topic modeling approach. J Med Internet Res 2020 Apr 28;22(4):e19118 [FREE Full text] [doi: 10.2196/19118] [Medline: 32302966]
11. Dong M, Cao X, Liang M, Li L, Liu G, Liang H. Understand research hotspots surrounding COVID-19 and other coronavirus infections using topic modeling. medRxiv Preprint posted online March 30, 2020. [doi: 10.1101/2020.03.26.20044164]

12. Stokes DC, Andy A, Guntuku SC, Ungar LH, Merchant RM. Public priorities and concerns regarding COVID-19 in an online discussion forum: longitudinal topic modeling. J Gen Intern Med 2020 Jul;35(7):2244-2247 [FREE Full text] [doi: 10.1007/s11606-020-05889-w] [Medline: 32399912]
13. Sha H, Al Hasan M, Mohler G, Brantingham PJ. Dynamic topic modeling of the COVID-19 Twitter narrative among U.S. governors and cabinet executives. arXiv Preprint posted online April 19, 2020.
14. Hosseini P, Hosseini P, Broniatowski DA. Content analysis of Persian/Farsi Tweets during COVID-19 pandemic in Iran using NLP. arXiv Preprint posted online May 17, 2020.
15. Sharma K, Seo S, Meng C, Rambhatla S, Liu Y. Covid-19 on social media: analyzing misinformation in Twitter conversations. arXiv Preprint posted online March 26, 2020.
16. Odlum M, Cho H, Broadwell P, Davis N, Patrao M, Schauer D, et al. Application of topic modeling to tweets as the Foundation for Health Disparity Research for COVID-19. Stud Health Technol Inform 2020 Jun 26;272:24-27 [FREE Full text] [doi: 10.3233/SHTI200484] [Medline: 32604591]
17. Wang X, Zou C, Xie Z, Li D. Public opinions towards COVID-19 in California and New York on Twitter. medRxiv Preprint posted online July 14, 2020. [doi: 10.1101/2020.07.12.20151936]
18. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
19. Ordun C, Purushotham S, Raff E. Exploratory analysis of Covid-19 tweets using topic modeling, UMAP, and DiGraphs. arXiv Preprint posted online May 6, 2020.
20. Hamilton W, Clark K, Leskovec J, Jurafsky D. Inducing domain-specific sentiment lexicons from unlabeled corpora. Proc Conf Empir Methods Nat Lang Process 2016 Nov;2016:595-605 [FREE Full text] [doi: 10.18653/v1/D16-1057] [Medline: 28660257]
21. Chen E, Lerman K, Ferrara E. Tracking social media discourse about the COVID-19 pandemic: development of a public coronavirus Twitter data set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
22. spacy-langdetect 0.1.2. The Python Package Index. URL: https://pypi.org/project/spacy-langdetect/ [accessed 2020-06-15]
23. tweet-preprocessor 0.6.0. The Python Package Index. URL: https://pypi.org/project/tweet-preprocessor/ [accessed 2020-06-15]
24. Honnibal M, Montani I. spacy 2: natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. Sentometrics Res 2017:1.
25. hyejuj / covid19. GitHub. URL: https://github.com/hyejuj/covid19 [accessed 2021-01-19]
26. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Machine Learning Res 2003:933-1022.
27. Pedregosa F, Varoquax G, Michel V, Thirion B, Grisel O, Blondel M, et al. Scikit-learn: machine learning in Python. J Machine Learning Res 2011:2825-2830.
28. Griffiths TL, Steyvers M. Finding scientific topics. Proc Natl Acad Sci U S A 2004 Apr 06;101 Suppl 1:5228-5235 [FREE Full text] [doi: 10.1073/pnas.0307752101] [Medline: 14872004]
29. Pereg O, Korat D, Wasserblat M, Mamou J, Dagan I. ABSApp: a portable weakly-supervised aspect-based sentiment extraction system. 2019 Presented at: 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations; November 2019; Hong Kong, China. [doi: 10.18653/v1/d19-3001]
30. Straus SE, Tetroe J, Graham I. Defining knowledge translation. CMAJ 2009 Aug 04;181(3-4):165-168 [FREE Full text] [doi: 10.1503/cmaj.081229] [Medline: 19620273]
31. Graham I, Logan J, Harrison M, Straus S, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
32. Yin H, Yang S, Li J. Detecting topic and sentiment dynamics due to COVID-19 pandemic using social media. arXiv Preprint posted online July 5, 2020.
33. Zhu H. Countering COVID-19-related anti-Chinese racism with translanguaged swearing on social media. Multilingua 2020;39(5):607-616. [doi: 10.1515/multi-2020-0093]
34. Chen L, Lyu H, Yang T, Wang Y, Luo J. In the eyes of the beholder: analyzing social media use of neutral and controversial terms for COVID-19. arXiv Preprint posted online April 21, 2020.
35. Croucher S, Nguyen T, Rahmani D. Prejudice toward Asian Americans in the Covid-19 pandemic: the effects of social media use in the United States. Front Commun 2020 Jun 12;5:1. [doi: 10.3389/fcomm.2020.00039]
36. Gore R, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to Twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
37. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]

Abbreviations

ABSA: aspect-based sentiment analysis
LDA: latent Dirichlet allocation

Edited by G Eysenbach; submitted 02.11.20; peer-reviewed by A Andy, R Gore; comments to author 08.12.20; revised version received 19.01.21; accepted 20.01.21; published 10.02.21.

Please cite as:
Jang H, Rempel E, Roth D, Carenini G, Janjua NZ
Tracking COVID-19 Discourse on Twitter in North America: Infodemiology Study Using Topic Modeling and Aspect-Based Sentiment Analysis
J Med Internet Res 2021;23(2):e25431
URL: http://www.jmir.org/2021/2/e25431/
doi:10.2196/25431
PMID:33497352

Original Paper

Measuring Public Reaction to Violence Against Doctors in China: Interrupted Time Series Analysis of Media Reports

Qian Yang[1], PhD; Ming Tai-Seale[2], PhD; Stephanie Liu[3], BA; Yi Shen[3], MA; Xiaobin Zhang[4], MA; Xiaohua Xiao[3], BA; Kejun Zhang[4], PhD

[1]Center for Health Policy Studies, School of Public Health and Department of Endocrinology, Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China

[2]Department of Family Medicine, School of Medicine, University of California San Diego, San Diego, CA, United States

[3]School of Public Health, Zhejiang University, Hangzhou, China

[4]College of Computer Science and Technology, Zhejiang University, Hangzhou, China

Corresponding Author:
Kejun Zhang, PhD
College of Computer Science and Technology
Zhejiang University
866 Yuhangtang Road
Hangzhou, 310058
China
Phone: 86 18758181126
Email: zhangkejun@zju.edu.cn

Abstract

Background: Violence against doctors in China is a serious problem that has attracted attention from both domestic and international media.

Objective: This study investigates readers' responses to media reports on violence against doctors to identify attitudes toward perpetrators and physicians and examine if such trends are influenced by national policies.

Methods: We searched 17 Chinese violence against doctors reports in international media sources from 2011 to 2020. We then tracked back the original reports and web crawled the 19,220 comments in China. To ascertain the possible turning point of public opinion, we searched violence against doctors–related policies from Tsinghua University ipolicy database from 2011 to 2020, and found 19 policies enacted by the Chinese central government aimed at alleviating the intense patient–physician relationship. We then conducted a series of interrupted time series analyses to examine the influence of these policies on public sentiment toward violence against doctors over time.

Results: The interrupted time series analysis (ITSA) showed that the change in public sentiment toward violence against doctors reports was temporally associated with government interventions. The declarations of 10 of the public policies were followed by increases in the proportion of online public opinion in support of doctors (average slope changes of 0.010, $P<.05$). A decline in the proportion of online public opinion that blamed doctors (average level change of –0.784, $P<.05$) followed the declaration of 3 policies.

Conclusions: The government's administrative interventions effectively shaped public opinion but only temporarily. Continued public policy interventions are needed to sustain the reduction of hostility toward medical doctors.

(J Med Internet Res 2021;23(2):e19651) doi:10.2196/19651

KEYWORDS

violence against doctors; government intervention; public opinion; patient–physician relationship

Introduction

With the improvement of Chinese people's living standard [1] and the attendant increase in personal wealth, individual Chinese citizens are paying more attention to their quality of life and health [2]. Health concerns have gradually become the focus of public discourse and are among the most popular topics on the internet in China. Concurrent with this trend, the plight of the strained doctor–patient relationship has become more evident, with violence against doctors, nurses, and other health care

providers in China on the rise [3]. Incidents of violence against doctors increased from 20.6 incidents per hospital in 2008 to 27.3 in 2012 [4]. An editorial in *The Lancet* in 2012 called this situation a "crisis" for the Chinese medical field [5].

In this period of mass media, online commentary and discussion have increasingly been proven as a useful instrument for surveying the moods and attitudes of the general public. The prevalence of the media and the internet for use in health communication has already become standard in regions such as China/Taiwan, where the media plays a significant role in settling medical disputes and dealing with medical institutions for patients [2]. For instance, the online community may share viewpoints on an individual physician or medical facility, which considerably shapes the image and evaluation of the physician or medical facility for the general reading public [6]. Moreover, those with media access that can further express their thoughts through media outlets can affect how news is portrayed and its content [7].

Public interest in the doctor–patient relationship and the widespread use of the internet have made it relatively easy for violence against doctors events to trigger a dispute of public opinion online. It has been suggested that the rise of online public discourse on violence against doctors may negatively affect the doctor–patient relationship and exacerbate the perception of social chaos and loss of control [8]. While many may agree that public opinions online are symbolic of challenges in real life, as reflected by online expressions of attitudes, views, and emotions, their impact reaches beyond the virtual world. Many studies have shown that online public opinion is relatively consistent in its intensity and continuity, influencing uncivilized network behaviors and social uncertainty [9,10].

To maintain social stability both online and offline, the Chinese government has made great efforts to regulate online content. For example, it is illegal to create or spread rumors on the internet in China as it may disrupt public order and social security [11]. In addition, the government has implemented a series of policies to increase the punishment for the crime of violence against doctors, and a government bureau, the National Radio and Television Administration, has published a series of articles to guide public opinion. Moreover, in July 2014, the National Health and Family Planning Commission, Ministry of Justice, Ministry of Finance, China Insurance Regulatory Commission, and National Administration of Traditional Chinese Medicine issued a vital policy—Opinions of the National Health and Family Planning Commission, the Ministry of Justice, the Ministry of Finance and Other Departments on Strengthening the Work of Medical Liability Insurance [12]—to establish a medical risk-sharing mechanism with medical liability insurance as the primary form and make this mechanism essential in medical dispute resolution and medical risk management. The Chinese Medical Association has issued a statement calling for system-wide reforms. In October 2013, the Ministry of Public Security of the People's Republic of China advised hospitals with over 2000 beds to hire "at least 100 security guards" [13]. However, increased security guards, metal detectors, and legal threats have been criticized for failing to deal with the underlying causes of the violence [14].

These policies have been attributed to some reduction in the acceleration of violence against doctors and calming of public opinion on doctor–patient conflicts [15]. Patients feel a lack of control due to the uncertainty of illness [16,17]. Many people require personal control over their lives and the surrounding environment, perhaps to avoid feeling the chaos and uncertainty in the world [18,19]. When people feel lower levels of personal control, they are more likely to show more support for government control [20].

However, whether such policies and interventions have made a practical positive impact on online public opinion has not been examined. One way to investigate this question is to compare the timeline of a turning point in public comments and the timing of the introduction of violence against doctors–related policies. Comments of the article instead of its content could provide more intuitive reflection of the patients side [21]. Accordingly, to gauge the overall public attitude toward doctors and the health care system in China, we endeavored to measure online reactions to mass media reports of violence against doctors.

Comments on violence against doctors incidents expose the attitude of the affected party groups to a certain extent. In recent years, an increasing number of violence against doctors events have affected people's perceptions of the country in a period of social transition, where intense emotions and perceptions of people and affected groups are also influenced by perceived sense of control [22], and government policy interventions can provide a compensatory sense of said control [20]. Therefore, to understand the emotional valence (sentimental valence) of public comment on violence against doctors episodes, it is meaningful to check whether national policy has affected the stereotypes associated with the doctor–patient bond from a time perspective.

To assess the effectiveness of government interventions in influencing online public opinion in China, we retrieved significant media reports of violence against doctors events and time-relative related government policies (eg, Criminal law or rules and regulations of the National Health Commission) in a policy database from 2011 to 2020 [12,15,23,24]. We documented public attitudes by coding comments on online media. Overall, we expected that these data would provide a better understanding of the general public who utilize social media and the internet, as well as insight into their attitudes toward health, perceived quality of health care services, and health care information needs. Thus, our research questions were (1) what trends in attitudes do the readers show with respect to the actions of both patients and physicians? (2) Are these trends influenced by the introduction and direction of national policies?

Methods

Sampling and Data Collection

To identify which events are internationally influential, we visited well-known media sites, including the Wall Street Journal, New York Times, BBC, Economist, Washington Post, Telegraph, Times, New Yorker, South China Morning Post, International Business Times, National Public Radio, and The

Atlantic and found a total of 17 medical violence incidents reported in China between October 19, 2011, and April 4, 2020. In total, there were 29 English reports. We then traced these incidents to Chinese website media and crawled (using a web robot to collect scripts at high speed) them for encoding and analysis to estimate online public attitudes toward doctors and patients (Multimedia Appendix 1).

We selected popular Chinese news media (N=6) from the New Media Influence Index Report [25], and considered their lasting influence from 2011, as shown in Table 1. We selected these 6 news media for 2 reasons. First, when we traced back the Chinese violence against doctors reports in international media sources, these media had a higher report rate than other media [26]; second, we referred to the media influence from 2011 to 2019 and, according to the Kantar China Media Impact Report 2016 and 2019, these 6 news media were more influential during these years [27].

We obtained 17 cases and their comments from the selected news media (see Multimedia Appendix 2 for details). As 2 of the 17 cases occurred in private hospitals, pseudonyms were applied to both private and public hospitals. There were 3 second-class A hospitals, 1 second-class B hospital, 2 top-class B hospitals, and 9 top-class A hospitals. Thus, the study covers different levels of hospitals, consistent with the distribution of hospital ownership status and hospital level in the total sample of doctor–patient disputes [28]. As the median number of comments was 400 on each incident, based on the total media coverage, if the count exceeded 400, we took a random sample of 400. The comments of each case are arranged in chronological order and divided into X/2n areas, and 2 comments were randomly extracted from each area. If the count was under 400, all coverage was included in the analyses. We then eliminated the repeated and unrelated comments. If the number of meaningless comments in the selected final comment sample exceeded 20%, these comments were removed. The number of comments required was then supplemented by the above sampling method. The sampling method is based on a real-world situation. We then collated and analyzed all the comments we collected. The final comment numbers for each case are listed in Table 2.

Table 1. All news media used to obtain sample comments.

News media	Reference
Tencent News	[29]
Sohu News	[30]
iFeng News	[31]
Sina News	[32]
Tianya Forum	[33]
NetEase News	[34]

Table 2. Number of network reviews for all cases.

Case	2011 (N=255)	2012 (N=1245)	2013 (N=178)	2014 (N=1473)	2015 (N=889)	2016 (N=1003)	2018 (N=924)	2019 (N=721)	2020 (N=521)	Total (N=7209)
C1	254	428	1	3	—[a]	—	—	—	—	696
C2	—	384	—	5	—	—	—	—	—	389
C3	1	433	38	2	—	—	—	—	—	474
C4	—	—	139	474	247	—	—	—	—	860
C5	—	—	—	400	2	—	—	—	—	402
C6	—	—	—	189	—	—	—	—	—	189
C7	—	—	—	—	200	—	—	—	—	200
C8	—	—	—	—	158	—	—	—	—	158
C9	—	—	—	—	282	2	—	—	—	284
C10	—	—	—	400	—	—	—	—	—	400
C11	—	—	—	—	—	400	—	—	—	400
C12	—	—	—	—	—	400	—	—	—	400
C13	—	—	—	—	—	201	—	—	—	201
C14	—	—	—	—	—	—	396	3	121	520
C15	—	—	—	—	—	—	528	—	—	528
C16	—	—	—	—	—	—	—	539	—	539
C17	—	—	—	—	—	—	—	179	400	579

[a]"—" represents no count of the comments in the *given* year.

Ethical Approval and Consent to Participate

Ethical approval was obtained from the School of Public Health, Zhejiang University. As open data obtained through web crawlers do not include any personal data, informed consent was not required (waived by the ethics committee), and the study involved minimal risk.

Coding

Procedure and Steps

A coding group comprising 4 preventive medicine students was trained to analyze the coverages. The coding forms outlined 9 categories developed a priori based on previous literature [35,36] (Table 3). The coding was performed in MS Excel.

Table 3. Coding categories.

Category	Meaning/Definition
Blame Big System	The comments express dissatisfaction, anger, or blame for the whole society or social atmosphere.
Blame Medical System	The comments express dissatisfaction, anger, or blame for health policy or the hospital system.
Blame Doctor	The comments express dissatisfaction, anger, or blame for the doctor or his/her behavior, other doctors, or their behavior in the case.
Blame Patient	The comments express dissatisfaction, anger, or blame for the patient or his/her behavior, other patients, or their behavior in the case.
Blame Other	The comments express dissatisfaction, anger, or blame for other things, such as the news media, the legal system.
Support Doctor	The comments express understanding, sympathy, or support for the doctor or his/her behavior, other patients, or their behavior in the case.
Support Patient	The comments express understanding, sympathy, or support for the patient or his/her behavior, other patients, or their behavior in the case.
Support Other	The comments express understanding, sympathy, or support for other things.
N/A	Comments are unrelated to cases or do not belong to any of the above categories.

At the beginning of the evaluation, we trained the 4 raters in the method of "independent coding–proofreading–third party opinion discussion" and then ensured that the interrater reliability reached 0.9 or above before they began to evaluate

the formal materials. All the formal comments assessments were completed according to the following steps.

Step 1

The coding team carried out training to define and unify the coding rules.

Step 2

Fifty sample comments were extracted, and each rater was asked to code these comments independently. We compared the coding results between raters to obtain the "interrater reliability" (interrater reliability = the number of entries with the same coding result/Total number of entries, ie, the consistency between all raters' coding results). The final results were unified according to the coding rules.

Step 3

If interrater reliability was less than 0.9, we repeated the operations in step 2 until it reached 0.9. The coding team then began to code formal samples.

Step 4

The raters read and coded all the comments independently, without interfering or discussing with each other. Each comment could belong in 1 category or more.

Step 5

We compared the coding results of each rater to find comments that were inconsistent in the coding results.

Step 6

We established a discussion group to reread all the comments and to determine the coding results of comments that were inconsistent in step 4. Finally, the coding team completed the coding results of all the comments.

Step 7

For the comments between 2017 and 2020, we used a sentimental dictionary developed from the code between 2011 and 2016. The sentiment was Normalization as [0,1].

Constructing a Sentiment Dictionary

We collected all the comments, marked each sentiment of each comment, and then segmented these comments into words, and removed some stop words to obtain the probability P_{ij} of each word appearing in each sentiment:

$$P_{ij}=(e_j|w_i) = x = [C(w_i,e_j)]/[C(e_j)]$$

where P_{ij} represents the probability that w_i represents sentiment e_j, and $C(w_i,e_j)$ represents the number of times that w_i appears in sentiment e_j.

Comment Sentiment Judgment

Thereafter, when judging new comments, we first segmented each comment and removed some stop words. According to the naïve Bayes principle, a certain sentiment of each comment is the accumulation of the sentiments of all participles of the comment. To eliminate the polarization problem, the calculation is multiplied by the value related to the length of each comment:

Finally, a threshold value is selected according to q_{kj} to judge the sentiment determination.

Policy Content Quantitative Analysis

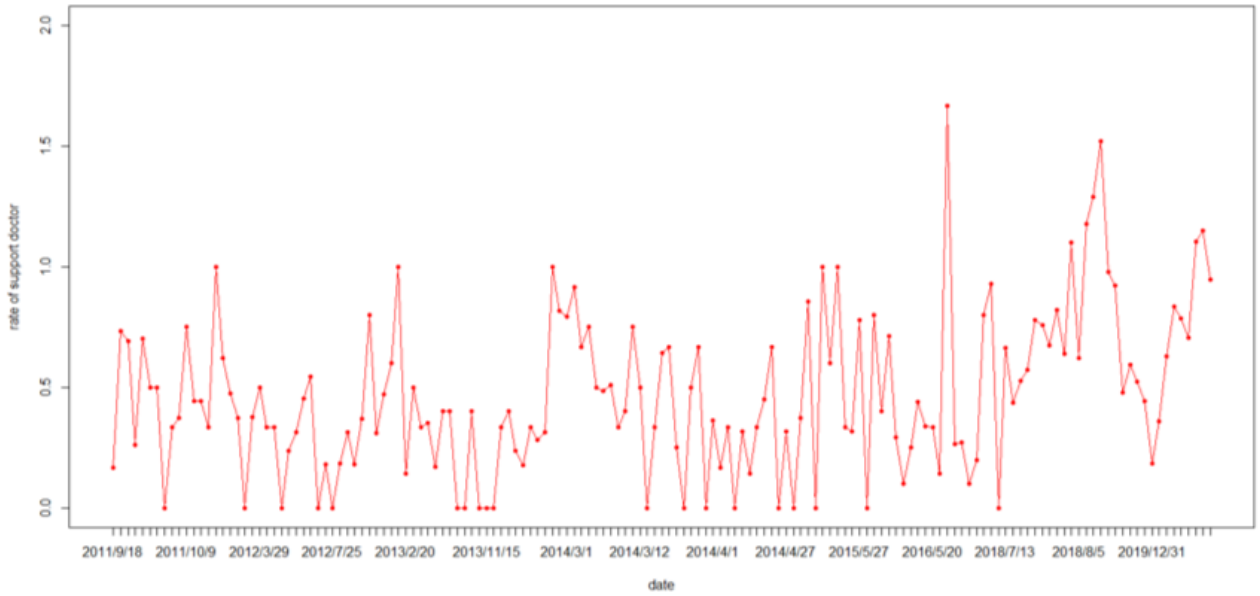
Policy content quantitative analysis is a semantic analysis method that combines quantitative and qualitative analysis of the content of policy literature. A proper policy database could display the essential information of the content of violence against doctors–related policy literature objectively [37].

This study used Tsinghua University's policy analysis system, "ipolicy," to retrieve policies. "ipolicy" is a unique characteristic analysis system, which can perform the specified analysis function according to users' needs at different levels. At present, there are approximately 1,200,000 policy texts from 1949 to 2019 (including all the organs of the State Council, other central bodies, the National People's Congress, and more than 34 provinces), which are still being updated [38].

In the retrieval module, the system provides various retrieval modes, and the content of the retrieval can be selected as full text or title retrieval. Advanced retrieval and retrieval in the results are also available to facilitate the user to find the text more accurately.

We used interrupted time series analysis (ITSA), a kind of longitudinal quasi-experimental design [39,40]. ITSA provides a segmented linear regression model, which could assess the effect of each central policy on changes in the "support doctor" and "blame doctor" online public opinion before the policy promulgation date and at 30 time points after the promulgation date. We combined some of the rates of sentimental categories to understand netizens' (Internet Citizens [41]) attitudes toward violence against doctors events. Because the meanings of "blame doctor" and "support patient" are similar, we combined the 2 categories of results into the new category of "blame doctor." Likewise, we combined the results of "blame patient" and "support doctor" into the new category of "support doctor." Originally, we crawled 19,220 comments. For the sake of ITSA, we averaged the sentiments of comments of the same date. For those with only 1 comment on each day, we framed adjected comments of 3-4 days and selected the second date to represent the time point. Finally, we obtained 151 time points representing the online public opinion change (Figure 1). As the policy promulgation date overlapped with the 151 points, we finally obtained 15 policy intervention dates for ITSA. The models quantify both a level and slope change following the intervention, while accounting for the autocorrelation of rates. Autoregressive integrated moving average models were used to adjust for residual autocorrelation. We prepared the data with MS Excel, version 2019, and completed all statistical analyses with R statistical software package, version 4.0.2 (R Foundation for Statistical Computing), using a Type 1 error rate of .05 as the threshold for statistical significance.

Figure 1. The rate of support doctors at the 151 time points from 2011–2020.



The regression model used to fit these data is straightforward:

$$outcome_{jt} = \beta_0 + \beta_1 \times time_t + \beta2 \times level_j + \beta3 \times trend_{jt} + \varepsilon_{jt}$$

In this specific example (using the variable names from Table 7), the model is:

$$sentiment\ rate_t = \beta_0 + \beta_1 \times time_t + \beta2 \times policy_j + \beta3 \times time\ after\ policy_{jt} + \varepsilon_{jt}$$

Results

Number and Percentage of Coding Categories

Table 4 lists the preliminary results of coding, including the number of comments belonging to each category in the years 2011-2020, and the percentage of total comments in this period.

Table 4. Number and percentage of coding categories.

Coding category	Year								
	2011 (N=255)	2012 (N=1245)	2013 (N=232)	2014 (N=1473)	2015 (N=889)	2016 (N=1003)	2018[a] (N=924)	2019[a] (N=721)	2020[a] (N=521)
Blame Big System, n (%)	30 (11.8)	132 (10.6)	25 (10.8)	174 (11.8)	58 (6.5)	74 (7.3)	(1.11)	(1.32)	(3.25)
Blame Medical System, n (%)	8 (3.1)	79 (6.3)	24 (10.3)	36 (2.4)	41 (4.6)	70 (6.9)	(0.34)	(1.2)	(2.19)
Blame Doctor, n (%)	85 (33.3)	312 (25.1)	91 (39.2)	191 (12.9)	183 (20.6)	133 (13.2)	(2.54)	(1.85)	(4.86)
Blame Patient, n (%)	68 (26.7)	167 (13.4)	24 (10.3)	362 (24.6)	161 (18.1)	207 (20.6)	(4.08)	(2.34)	(12.71)
Blame Other, n (%)	4 (1.6)	13 (1.0)	3 (1.3)	47 (3.2)	24 (2.7)	24 (2.3)	(0.0)	(0.56)	(1.92)
Support Doctor, n (%)	65 (25.5)	210 (16.9)	46 (19.8)	264 (17.9)	113 (12.7)	340 (33.9)	(2)	(1.95)	(3.83)
Support Patient, n (%)	60 (23.5)	71 (5.7)	15 (6.5)	19 (1.3)	44 (4.9)	8 (0.8)	(0.53)	(0.92)	(2.52)
Support Other, n (%)	0 (0.0)	1 (0.1)	1 (0.4)	1 (0.1)	2 (0.2)	0 (0.0)	0 (0.0)	(0.55)	(1.92)
N/A, n (%)	25 (9.8)	443 (35.6)	47 (20.3)	562 (38.2)	391 (43.9)	461 (45.9)	(0.82)	(1.77)	(3.72)

[a]Because the rate of 2017-2020 was from data mining, n is not available.

As shown in Table 4, the proportion of “blame big system” is more stable than other categories; “blame doctor” peaked at 39.2% (91/232) in 2013, but decreased to 12.97% (191/1473) in 2014 and 4.86% in 2020; “support doctor” decreased from 2011 to 2012, but rose significantly from 2015 to 2020; “blame patients” reached the minimum of 10.3% (24/232) in 2013 and the maximum of 24.58% (362/1473) in 2014; “support patient” showed a downward trend, from the overall in 2020, when it was only 2.52% (Figures 2 and 3).

Figure 2. "Support doctor" proportion of public online opinion over time. Trend lines show slope change after lagged intervention date. Vertical dash line shows lagged intervention date of policy 2, Dec 20, 2013.

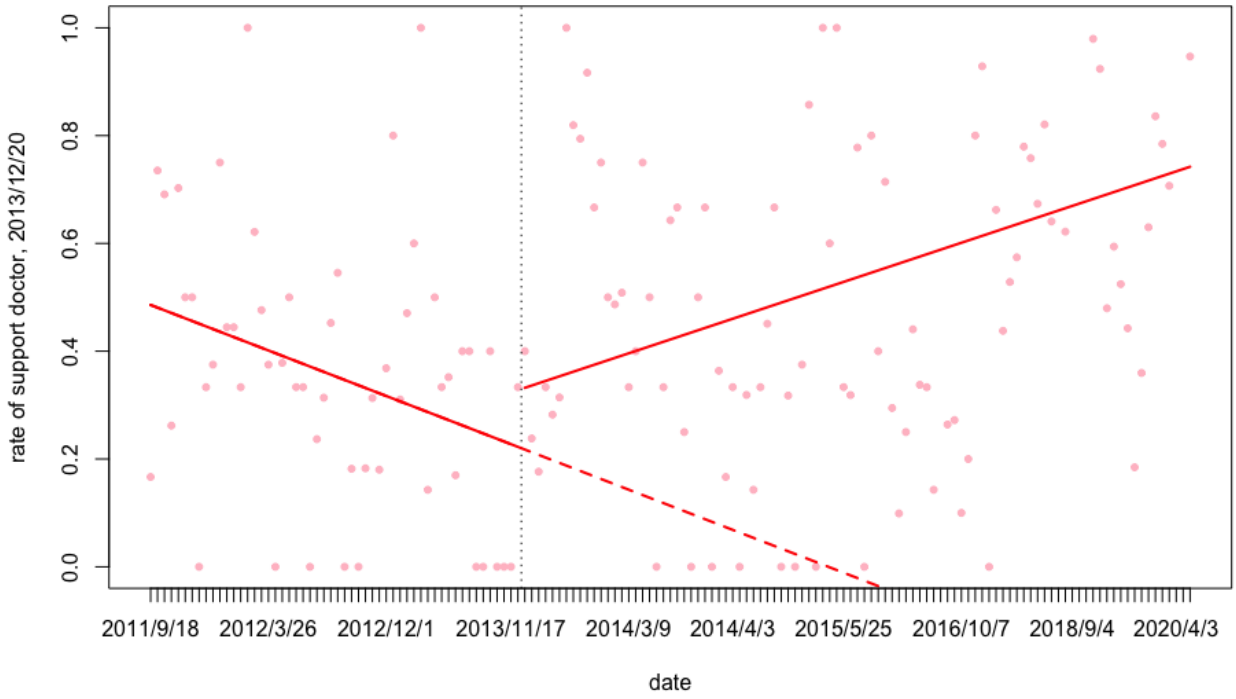
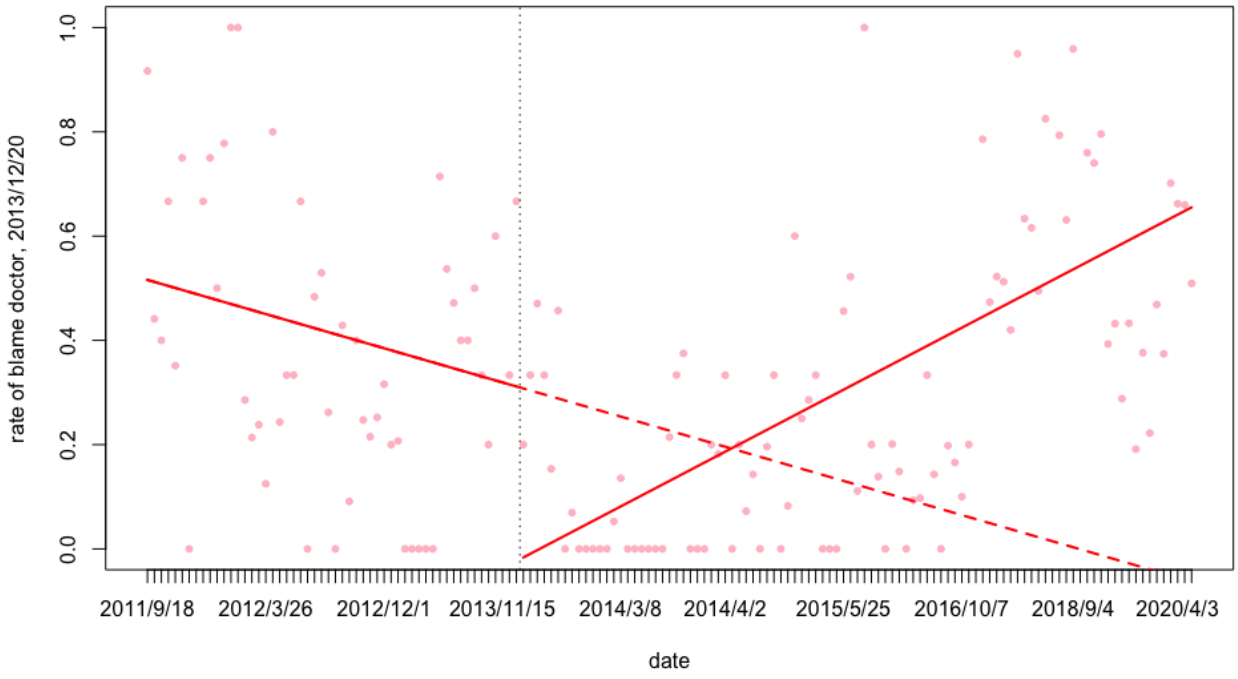


Figure 3. "Blame doctor" proportion of public online opinion over time. Trend lines show slope change after lagged intervention date. Vertical dash line shows lagged intervention date of policy 2, Dec 20, 2013.



Retrieval Results From the "ipolicy" System

The keywords we used to retrieve related policies were "doctor-patient," "medical," "violence," "medical violence," and "violence against doctors." We then obtained related policies and regulations from 2011 to 2016 in China (Tables 5 and 6).

Table 5. Retrieval results from the "ipolicy" system.

Keywords	Results
Medical violence (医疗暴力)	Not related to the topic
Doctor-patient (医患)	Not related to the topic
Medical + violence (医疗或暴力)	Not related to the topic
Hurt doctors (伤医)	Two related entries in total
Violence against doctors (医暴)	Ninety-seven related entries in total

Table 6. Policy entries and netizens' attitude trend.

Year	Policy entries, n			Netizens' attitude trend, n	
	Total	Central policies	Local policies	Blame doctor	Support doctor
2011	12	0	19	—[a]	—
2012	24	11	12	↓[b]	↓
2013	27	16	11	↑[c]	—
2014	27	12	13	↓	↑
2015	4	4	0	↑	↓
2016	2	2	0	↓	↑
2017	3	3	0	↓	↓
2018	0	0	0	↓	↓
2019	0	0	0	—[d]	—
Sum	99	48	55		

[a]No change in attitude.

[b]Decrease in ratio of related attitude.

[c]Increase in ratio of related attitude compared with last year.

[d]Not available.

We browsed these 112 policies, and 19 policies were promulgated by the central government (Multimedia Appendix 3), of which the most relevant central policy documents are as follows:

1. Severe punishment of any illegal activities that would harm the safety of doctors or patients. We will severely punish illegal activities that disrupt regular medical order and improve the ability to solve medical disputes [15].
2. Keep improving awareness of and fight against illegal medical activities and crimes [23].
3. Resolutely combat crimes involving hospitals to maintain order and to punish mob behavior, etc. [24].
4. Establish a medical risk-sharing mechanism with medical liability insurance as the main form, and have this mechanism play an important role in medical dispute resolution and medical risk management [12].

As shown in Tables 7 and 8, for the first central policy (promulgated on October 26, 2012), at baseline (before the policy was promulgated), there was neither a significant level nor trend change in the rate of "support doctor." However, there was a drop of 2% per time point in the "blame doctor" trend (Figures 2 and 3).

There was an increase in the "support doctor" trend after each policy promulgation (average trend changes of 0.011 of the policies 2-12, $P<.05$), but most of the slope change was not significant. At almost every time point in the policy promulgation, there was a decline in the proportion of "blame doctor" (average level drop of –0.784 of the policies 2-4, $P<.05$); however, there was also a trend change lag (average slope change of 0.018 of the policies 2-12, $P<.05$).

Table 7. Support doctor regression parameters at 15 policy time points.

Policy	Policy date	β1	*P*	β2	*P*	β3	*P*
1	October 26, 2012	–.009	.235	.035	.825	.013	.097
2	December 20, 2013	–.005	.124	.106	.391	.009	.010
3	January 29, 2014	–.005	.176	.126	.365	.009	.023
4	March 26, 2014	–.005	<.001	–.165	.034	.016	<.001
5	April 22, 2014	.001	.789	–.125	.169	.010	<.001
6	May 8, 2014	.001	.881	–.041	.679	.010	<.001
7	July 9, 2014	.001	.717	.005	.958	.010	<.001
8	September 4, 2014	.001	.943	–.064	.519	.011	<.001
9	October 27, 2014	.001	.660	–.104	.318	.012	<.001
10	June 24, 2015	.001	.688	.007	.929	.011	<.001
11	March 24, 2016	.001	.646	–.041	.708	.013	<.001
12	May 13, 2016	.001	.863	.067	.553	.011	.007
13	June 26, 2017	.001	<.001	.123	.272	.013	.066
14	July 31, 2018	.001	.081	.441	.025	.005	.781
15	March 5, 2019	.002	.109	.120	.587	.006	.766

Table 8. Blame doctor regression parameters at 15 policy time points.

Policy	Policy date	β1	*P*	β2	*P*	β3	*P*
1	October 26, 2012	–.017	.057	–.038	.8316	.020	.031
2	December 20, 2013	–.004	.2478	–.335	.007	.011	.002
3	January 29, 2014	–.005	.1259	–.277	.046	.012	.002
4	March 26, 2014	–.004	<.001	–.172	<.001	.015	<.001
5	April 22, 2014	–.005	<.001	.025	.4097	.016	<.001
6	May 8, 2014	–.005	<.001	.103	.001	.016	<.001
7	July 9, 2014	–.005	<.001	.128	<.001	.015	<.001
8	September 4, 2014	–.005	<.001	.151	<.001	.015	<.001
9	October 27, 2014	–.005	<.001	.256	<.001	.013	<.001
10	June 24, 2015	–.004	<.001	.332	<.001	.011	<.001
11	March 24, 2016	–.004	<.001	.435	<.001	.008	<.001
12	May 13, 2016	–.004	<.001	.398	<.001	.010	.018
13	June 26, 2017	–.002	.062	.358	.038	.004	.646
14	July 31, 2018	.000	.898	.120	.562	–.002	.937
15	March 5, 2019	.001	.716	–.251	.228	.008	.7322

Discussion

Principal Findings

In 2017, discussion in *The Lancet* suggested that recommendations to rebuild patient–physician trust in China should not only focus on the physician's side but also on the patient's side [42,43]. Recently, particularly during COVID-19, voices to stop attacks against health care personnel have been raised worldwide [44-46]. All the health professional associations, societies, and organizations from all specialties and disciplines should unite to protect staff, while the government should play a major role by enacting strong and appropriate policies [47].

The current research has found that the 14 violence against doctors–related policies promulgated by the Chinese central government are associated with the sentiment expressed in online public opinions. The rate of online comments supporting doctors tends to increase significantly (see Table 7 for significant values) after almost each policy promulgation, while the rate of blaming doctors declined at each policy promulgation time point. However, the trend of blaming doctors then rises later

after every policy is enacted. The overall declining trend of blaming doctors (39.22% in 2013 versus 4.86% in 2020) and the increased ITS trend for each policy time point imply that policy effects are only temporary. Particularly, in recent years (2017-2020), there have been few polices, and the ITS trends are not significant. It is time to reflect on the effects of policies.

The recent discussions regarding the relationship between violence against doctors events and health care insurance are very lively. In 2018, China's Congress promulgated the Social Insurance Law of the People's Republic of China (2018 Amendment) [48]. Although China's medical security system is continually improving, there are still some problems. Furthermore, the severe current aging of our society and young people's increasing work pressure have led to an increase in the incidence of some major diseases. For some significant diseases, many drugs are not included in the medical insurance coverage, and patients must bear most or all of the medical expenses alone. Patients are likely to experience significant imbalances, and it is easy to dissolve this inconsistency. The balance is blamed on the medical institution and the most directly contacted population—medical staff [49-51]. Our study has provided this new patient viewpoint from the health and science education for the public. In this study, we focused on netizens' comments on major violence against doctors events and their attitudes toward doctors, patients, and the government as direct evidence to understand public opinions and the state of the physician–patient relationship. We found a significant relationship (see Tables 7 and 8 for significant *P*-values) between the time of implementation of government health reform policies and the shift of public opinion from "blame doctor" to "support doctor." In 2011 and 2012, the proportion of "blame doctor" and "support doctor" opinions were relatively balanced. However, in 2013, the proportions began to shift: the proportion of "blame doctor" rose while "support doctor" dropped. Because of this differentiation, the government had to adopt some policies. By the end of 2013, the public mood had eased up after the implementation of related health policies. In 2014, the number of related health policies peaked but then decreased in 2015. Thus, the proportion of "support doctor" opinions fell in 2015. Again, the situation improved after the enactment of central regulations in 2016 (Figures 2 and 3).

From the "ipolicy" system, we retrieved related policies and regulations from 2011 to 2016 in China and found the most relevant central policies to be those decreeing stricter enforcement and severe punishment for those endangering the safety of doctors and other medical staff [27]. Furthermore, in 2014 and 2016, further policies showed more awareness of fighting illegal medical activities [23] and concerns for resolving mob-related criminal behavior at medical hospitals [24]. Before these enactments, the enforcement of punishment for violence against doctors and the level of security for hospital staff were weak, allowing doctors and other medical staff to become victims [52].

Nonetheless, it is the role of the government to enforce proper regulations to prevent danger to medical staff. Our results show that the only interaction between attitude and turning point occurred from 2013 to 2014, which is connected with relevant policies between 2014 and 2016 contributing to a significant drop in citizens' "blame" of doctors (see Table 8 for significant *P* values). These efforts show that the Chinese government has heightened vigilance toward the problems displayed by increased violence against doctors incidents and more order in controlling the harms caused by such incidents. However, their efforts remain slightly below the US approach, suggesting that the Chinese government should continue to adopt more control and management roles in medical emergencies. According to the Chinese government, "Following the Occupational Safety and Health Act of 1970, 26 US states and two US territories require employers to develop detailed violence prevention programs. The Occupational Safety and Health Administration (OSHA) have also increased the number of visits to medical facilities, from 11 in 2010 to 86 in 2014. The American Government Accountability Office (GAO) gave a specific recommendation in its 2016 Work Report. Employers should be punished if they allow employees to be exposed to potential workplace violence, especially if they have a history of violence, assault or physical assault; provide additional information to inspectors in specific court summon cases to help develop a system of summonses; follow up confirmation of the need for further measures" [42,43,53].

In analyzing the comments rather than the news articles themselves, we provided a more direct measure of the public's perspective and attitude toward the health care sphere and their stance on the patient–physician bond. The data we collected on these comments can further be used as a potential indicator of health care quality, as few reliable measures exist. Wang et al [54] utilized medical malpractice incidents as a probable marker of health care quality, but a more robust and direct measure is needed to probe the difficulties experienced by the Chinese medical field. We analyzed the comments based on the variable of time, so we can see changes and patterns through time, compared to previous studies, which examined more qualitative methods to seek reliable measures of the patient–doctor relationship [55,56]. Comments made by the general public can reveal their emotions and opinions, as well as their perceptions and viewpoints on the shortcomings of the health care field and expose needed reforms for the improvement of trust in medical professionals.

Our study also contributes to the qualitative study of feelings of mistrust in the physician–patient relationship in China. A survey of 107 health care physicians in Guangdong province, China, reported that physicians are caught in a "vicious cycle of mistrust" due to various contributing factors of work overload, conflicts of interest, unreasonable requests of patients, and more [57]. Restoring public trust in their physicians requires a strong foundation supported by government regulations for effective delivery and more transparent methods of health care quality assessments.

The media play an essential role in contributing to overall public attitudes toward the physician–patient relationship and overall health care system when it documents violence against doctors events for citizens. Both print and electronic media have a responsibility not to sensationalize the news. Health care workers and officials have cited negative media coverage and sensationalism of violence against doctors events as a prime reason contributing to the deteriorating doctor–patient

relationship and mistrust in the medical field [58]. The practice of medicine is highly complex. Diagnosis of a patient is an essentially hypothetico-deductive process, and with new evidence presented through investigations and knowledge, this process continues to be refined. However, whatever the diagnosis may be, patient management generally considers such uncertainties and treatment continues. One of the purposes of this paper has been to analyze whether the government's policies can affect public opinion on doctor–patient issues, mainly regarding the policies issued by the government on how to report violence against doctors events to influence public opinion. We expect that government policy could intervene in doctor–patient relationships through medical information and media, and our data demonstrate this. However, we do not expect that the decrease in netizens' blame of doctors and the increase in support for doctors will immediately end the trend of violent acts against doctors. Another purpose of our study was to understand the current trends in the medical network for violence against doctors and hope to find a breakthrough to ease public opinion. Although the media is not the initiator of such medical violence incidents, they are a vital participant in the aftermath, playing the critical role of disseminator of information. Thus, information authenticity and other aspects need to be considered.

In China, the internet environment is managed and controlled by the government, which pays attention to the impact of the network on society. If news media views drive the mood of internet users, then internet users' irrational speech will increase, also causing more extreme emotions and comments (broken window effect). Media reports tend to play an essential role in driving social norms, which influence people's behavior and outlooks on topics [57]. To prevent the news of medical violence harming public opinion, network administrators may need to carry out monitoring measures. In our study, we observed numerous emotionally charged comments, which indicate that these news media network comments have not been filtered by management; otherwise, there would be fewer "blame-style" comments. However, measuring the extent of comment monitoring of online news platforms is beyond the scope of our study. When an outlet is provided for netizens to discuss current events, express their views, and vent their emotions, it should guarantee as much free speech as possible, but it is unclear whether the same should apply in terms of medical violence.

China's intentional homicide has increased in our statistical data years—eg, from 10,285 in 2013 to 17,946 in 2015 [59]. The killing of a doctor is intentional homicide and with the increase in this crime, there has been an increase in the number of doctors killed. However, this does not mean that the relationship between doctors and patients has deteriorated further. The current research focuses on violence against doctors, not just homicides. According to statistics on other similar topics, the incidence of violent medical injuries in China has declined since 2015 [60].

Our study has several limitations. Our scope is limited to the violence against doctors events we selected for the study and to comments referring to those incidents, but we did not cover all types of medical incidents. Furthermore, the sample size, timeline, and number of comments are limited for objective reasons. Hence, we expect further research to be conducted to improve the study of people's emotional valence in violence against doctors events. In future research, we hope to study the key factors influencing people's emotions in violence against doctors events through randomized controlled trials. Future studies could also compare violence against doctors events across hospital types, for example, comparing violence against doctors events in private versus public hospitals. Examinations of violence against doctors events across departments within hospitals could also offer information on variations across specialties (eg, obstetrics, pediatrics, and other departments). Regional variations could be another area of inquiry. We seek to provide practical evidence and interventions to reduce the incidence of violence against doctors events and improve the doctor–patient relationship in China.

The causes of the rise and severity of violence against doctors in China are numerous and can be traced back to the systemic roots of inadequate investment in the health care system with subpar training and salaries for physicians, leading to higher chances of medical errors, administrative exploitation, and ineffective communication between patients and medical staff. Other factors may be cultural and societal, such as unfavorable and sensationalized news articles about doctors, the public's limited knowledge of medicine, unrealistic expectations from doctors and treatments, and detrimental out-of-pocket expenses for patients. As such, the impact of violence against doctors is of great concern and should not be taken lightly for the future of Chinese medicine. Further studies highlighting the role of the government and media should be conducted to improve the physician–patient relationship.

Conclusions

By matching the policy documents for medical violence with the online commentary attitudes, we found that the state's administrative intervention effectively guided public opinion. Notably, these findings provide scientific evidence and support for the government's role in intervention and prevention of both cyber and physical violence.

During the COVID-19 pandemic, many countries have been advocating home isolation, leading to a major increase in people's reliance on the internet for social support and health care information. During the pandemic, a harmonious patient–physician relationship has become an important guarantee for countries in responding to the epidemic. The results of this study show that the government's reasonable and effective policy guidance for violence against doctors has a significant impact (see Tables 7 and 8 for significant values) on the public's comments on online media reports, which in turn contribute to the normal medical order as well as epidemic prevention and control.

Acknowledgments

We thank Peng Wu, Jinrui Gao, Renjie Miao, and Zhen Yuan for their coding, and Mengfei Yu for her combining of the coding data. Hao Lei for his comments on statistics; and two anonymous reviewers for their guidance.

This study is supported by the National Natural Science Foundation of China (71974170), and Soft Science Project of Science Technology Department of Zhejiang Province of China (2019C35025). Dr. Ming Tai-Seale received no financial support for her work on this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comments acquisition.
[DOCX File , 16 KB - jmir_v23i2e19651_app1.docx]

Multimedia Appendix 2
Details of the VAD cases reported by international media from 2011–2020.
[DOCX File , 18 KB - jmir_v23i2e19651_app2.docx]

Multimedia Appendix 3
Nineteen policies promulgated by the Chinese central government.
[DOCX File , 18 KB - jmir_v23i2e19651_app3.docx]

References

1. Liew L. The Chinese Economy in Transition: From Plan to Market. Cheltenham (England) and Brookfield (Vermont): Edward Elgar Publishing Limited; 1997.
2. Xi J. Full text of Xi Jinping's report at 19th CPC National Congress. China Daily. 2017 Oct 18. URL: http://www.chinadaily.com.cn/china/19thcpcnationalcongress/2017-11/04/content_34115212.htm [accessed 2021-02-03]
3. Hesketh T, Wu D, Mao L, Ma N. Violence against doctors in China. BMJ 2012 Sep 07;345(sep07 1):e5730. [doi: 10.1136/bmj.e5730] [Medline: 22960376]
4. Pan J, Liu D, Ali S. Patient dissatisfaction in China: What matters. Soc Sci Med 2015 Oct;143:145-153. [doi: 10.1016/j.socscimed.2015.08.051] [Medline: 26356826]
5. The Lancet. Ending violence against doctors in China. The Lancet 2012 May;379(9828):1764. [doi: 10.1016/s0140-6736(12)60729-6]
6. Lupton D, McLean J. Representing doctors: discourses and images in the Australian press. Soc Sci Med 1998 Apr;46(8):947-958. [doi: 10.1016/s0277-9536(97)10013-2] [Medline: 9579747]
7. Happer C, Philo G. The Role of the Media in the Construction of Public Belief and Social Change. J. Soc. Polit. Psych 2013 Dec 16;1(1):321-336. [doi: 10.5964/jspp.v1i1.96]
8. Rippon TJ. Aggression and violence in health care professions. J Adv Nurs 2000 Feb;31(2):452-460. [doi: 10.1046/j.1365-2648.2000.01284.x]
9. Mouhssine E, Khalid C. Social Big Data Mining Framework for extremist content detection in Social Networks. New York: IEEE; 2018:978-971.
10. Chen H. Sentimentaffect analysis of dark web forums: Measuring radicalization on the internet. In: Yang CC, Chen WH, Hsu WL, Wu TC, Zeng D. eds. New York: IEEE; 2008:104-109.
11. Ministry of Justice of the People's Republic of China. Criminal Law of the People's Republic of China. 2020 Feb 14. URL: http://www.moj.gov.cn/subject/content/2020-02/14/1449_3241661.html [accessed 2021-02-05]
12. National Health and Family Planning Commission. National Opinions of the National Health and Family Planning Commission, the Ministry of Justice, the Ministry of Finance, and their Departments on Strengthening the Work of Medical Liability Insurance. 2015 Apr 1. URL: http://www.scio.gov.cn/32344/32345/32347/20150123/xgzc32642/Document/1397803/1397803.htm [accessed 2021-02-05]
13. National Health Commission of the People's Republic of China. A circular on the issuance of a special action plan to maintain medical order and crack down on crimes related to medical treatment. 2013 Dec 20. URL: http://www.nhc.gov.cn/yzygj/s3589/201312/46ce2b7c0b974391b8f8bf308f2784f8.shtml?from=groupmessage [accessed 2020-02-05]
14. Tang J, Kuang C. Can Yinao be stopped by the security hospital? Today's Morning Post (in Chinese). 2020 Apr 13. URL: http://news.sina.com.cn/c/2013-10-25/023028525228.shtml [accessed 2021-02-05]

15. National Health Commission of the People's Republic of China. Notice on the issuance of a special action plan to maintain medical order and crack down on medical crimes [Article 43]. 2020 Jun 26. URL: http://www.nhc.gov.cn/yzygj/s3589/201312/46ce2b7c0b974391b8f8bf308f2784f8.shtml [accessed 2021-02-05]
16. Baker P. Illness meanings and perceptions of control and uncertainty in women with breast cancer. In: Master Thesis. Newfoundland,CA: Memorial University of Newfoundland; 1994:97-102.
17. Ko N, Hsu S. Informational needs, health locus of control and uncertainty among women hospitalized with gynecological diseases. Chang Gung Med J 2005 Aug;28(8):559-566. [Medline: 16265846]
18. Leotti LA, Iyengar SS, Ochsner KN. Born to choose: the origins and value of the need for control. Trends Cogn Sci 2010 Oct;14(10):457-463 [FREE Full text] [doi: 10.1016/j.tics.2010.08.001] [Medline: 20817592]
19. Averill JR. Personal control over aversive stimuli and its relationship to stress. Psychological Bulletin 1973;80(4):286-303. [doi: 10.1037/h0034845]
20. Kay AC, Gaucher D, Napier JL, Callan MJ, Laurin K. God and the government: Testing a compensatory control mechanism for the support of external systems. Journal of Personality and Social Psychology 2008;95(1):18-35. [doi: 10.1037/0022-3514.95.1.18]
21. Zhang W, Deng Z, Hong Z, Evans R, Ma J, Zhang H. Unhappy Patients Are Not Alike: Content Analysis of the Negative Comments from China's Good Doctor Website. J Med Internet Res 2018 Jan 25;20(1):e35 [FREE Full text] [doi: 10.2196/jmir.8223] [Medline: 29371176]
22. Yang Q, Pan J. Control under times of uncertainty: the relationship between hospital competition and physician-patient disputes. Int J Equity Health 2017 Nov 28;16(1):1-12. [doi: 10.1186/s12939-017-0701-0]
23. National Health Commission of the People's Republic of China. The Comprehensive Treatment of Social Management Committee. Notice of the Ministry of Public Security on deepening the establishment of ping an hospital to further crack down on doctor-related crimes and maintain normal medical order. 2014 Sep 5. URL: http://www.nhc.gov.cn/yzygj/s3589/201409/b89257bb411d4ba58139a40c9b2f8872.shtml [accessed 2021-02-05]
24. National Health Commission of the People's Republic of China, The Comprehensive Treatment of Social Management Committee , The Ministry of Public Security of the People's Republic of China. Notice from the Ministry of Justice on further efforts to maintain medical order. 2016 Mar 30. URL: http://www.nhc.gov.cn/yzygj/s7658/201603/821b3e6e99e945088605161df0a856a0.shtml [accessed 2021-02-05]
25. Liu Z. New Media Influence Index Report. Beijing: China Social Sciences Press; 2019.
26. Kantar. China's Social Media Influence Rankings 2016. 2016 Jan 28. URL: http://datanews.caixin.com/2016/kantar16/index.html [accessed 2021-02-05]
27. Kanter. Kantar China Media Impact Report. 2016. URL: http://www.199it.com/archives/435396.html [accessed 2020-06-26]
28. Wang L, Jiang X, Zhang S, He L, Peng L. Statistical analysis of medical damage liability dispute cases in China. Chin J Health Stats 2016;33(6):1051-1053.
29. Tencent News. URL: http://news.qq.com/ [accessed 2021-02-06]
30. Sohu News. URL: http://news.sohu.com [accessed 2021-02-06]
31. iFeng News. URL: http://news.ifeng.com [accessed 2021-02-06]
32. Sina News. URL: http://news.sina.com.cn [accessed 2021-02-06]
33. Tianya Forum. URL: http://bbs.tianya.cn [accessed 2021-02-06]
34. NetEase News. URL: http://news.163.com [accessed 2021-02-06]
35. Davey K, Ravishankar V, Mehta N, Ahluwalia T, Blanchard J, Smith J, et al. A qualitative study of workplace violence among healthcare providers in emergency departments in India. Int J Emerg Med 2020;13(1):1-9. [doi: 10.1186/s12245-020-00290-0]
36. Vora P. India's hospital assaults: patients blame doctors while doctors blame the broken health system. Scroll.in. 2017 Mar 21. URL: https://scroll.in/pulse/832349/hospital-violence-patients-blame-doctors-while-doctors-blame-the-broken-health-system [accessed 2021-02-05]
37. Cui H, Tao R, Jian Z. Quantitative Research on Policy Literature: The New Direction of Public Policy Research. J Public Manag 2015;2:158-115.
38. Ipolicy PAS. Center for Government Literature. 2017. URL: http://39.105.58.246/ipolicy [accessed 2021-02-05]
39. Giuliani ME, Liu G, Xu W, Dirlea M, Selby P, Papadakos J, et al. Implementation of a Novel Electronic Patient-Directed Smoking Cessation Platform for Cancer Patients: Interrupted Time Series Analysis. J Med Internet Res 2019 Apr 09;21(4):e11735 [FREE Full text] [doi: 10.2196/11735] [Medline: 30964445]
40. Choi S, Yuen HM, Annan R, Pickup T, Pulman A, Monroy-Valle M, et al. Effectiveness of the Malnutrition eLearning Course for Global Capacity Building in the Management of Malnutrition: Cross-Country Interrupted Time-Series Study. J Med Internet Res 2018 Oct 03;20(10):e10396 [FREE Full text] [doi: 10.2196/10396] [Medline: 30282620]
41. Xu Y. Understanding netizen discourse in China: formation, genres, and values. China Media Res 2012;8(1):15-24.
42. Lyu Z, Wu S, Cai Z, Guan X. Patient–physician trust in China: health education for the public. The Lancet 2016 Dec 17;388(10063):2991. [doi: 10.1016/S0140-6736(16)32521-1]
43. Tucker JD, Wong B, Nie J, Kleinman A. Rebuilding patient–physician trust in China. The Lancet 2016 Aug 20;388(10046):755. [doi: 10.1016/s0140-6736(16)31362-9]

44. McKay D, Heisler M, Mishori R, Catton H, Kloiber O. Attacks against health-care personnel must stop, especially as the world fights COVID-19. The Lancet 2020 Jun;395(10239):1743-1745. [doi: 10.1016/s0140-6736(20)31191-0]
45. Dzau VJ, Kirch D, Nasca T. Preventing a Parallel Pandemic — A National Strategy to Protect Clinicians' Well-Being. N Engl J Med 2020 May 13;383:513-515. [doi: 10.1056/nejmp2011027]
46. Zhang Z, Liu S, Xiang M, Li S, Zhao D, Huang C. Protecting healthcare personnel from 2019-nCoV infection risks: lessons and suggestions. Frontiers of Medicine 2020;13:229-231. [doi: 10.1007/s11684-020-0765-x]
47. Zhang X, Guo X, Lai K, Yi W. How does online interactional unfairness matter for patient–doctor relationship quality in online health consultation? The contingencies of professional seniority and disease severity. European Journal of Information Systems 2018 Nov 27;28(3):336-354. [doi: 10.1080/0960085x.2018.1547354]
48. Standing Committee of National People's Congress. Social Insurance Law of the People's Republic of China (2018 Amended). 2018 Dec 19. URL: http://www.lawinfochina.com/display.aspx?id=29710&lib=law [accessed 2021-02-05]
49. Zhu K, Zhang L, Yuan S, Zhang X, Zhang Z. Health financing and integration of urban and rural residents' basic medical insurance systems in China. Int J Equity Health 2017 Nov 7;16(1):1-8. [doi: 10.1186/s12939-017-0690-z]
50. Zhang J, Cai Y. Medical disputes and mediation in China: Government and responsibility shifting. China Information 2018 Nov 11;33(3):350-371. [doi: 10.1177/0920203x18811038]
51. An P, Ye Y, Li Q, Liu B, Lian K, Yin J, et al. Medical disputes in relation to prenatal ultrasound in China. Ultrasound Obstet Gynecol 2020 Jul;56(1):11-14. [doi: 10.1002/uog.22020]
52. Peng W, Ding G, Tang Q, Xu L. Continuing violence against medical personnel in China: A flagrant violation of Chinese law. Bioscience trends 2016;10(3):240-243. [doi: 10.5582/bst.2016.01094]
53. Wyatt R, Anderson-Drevs K, Van Male LM. Workplace Violence in Health Care. JAMA 2016 Sep 13;316(10):1037. [doi: 10.1001/jama.2016.10384]
54. Wang Z, Li N, Jiang M, Dear K, Hsieh C. Records of medical malpractice litigation: a potential indicator of health-care quality in China. Bulletin of the World Health Organization 2017 Mar 13;95(6):430-436. [doi: 10.2471/BLT.16.179143]
55. Liu Z, Zhang Y, Asante JO, Huang Y, Wang X, Chen L. Characteristics of medical disputes arising from dental practice in Guangzhou, China: an observational study. BMJ Open 2018 Feb 08;8(2):e018738. [doi: 10.1136/bmjopen-2017-018738]
56. Nie J, Li L, Gillett G, Tucker JD, Kleinman A. The crisis of patient-physician trust and bioethics: lessons and inspirations from China. Developing World Bioeth 2017 Sep 18;18(1):56-64. [doi: 10.1111/dewb.12169]
57. Wilson JQ, George LK. Broken Windows: the police and neighborhood safety. The Atlantic Monthly 1982;29:37.
58. Nie J, Cheng Y, Zou X, Gong N, Tucker JD, Wong B, et al. The vicious circle of patient-physician mistrust in China: health professionals' perspectives, institutional conflict of interest, and building trust through medical professionalism. Developing World Bioeth 2017 Sep 18;18(1):26-36. [doi: 10.1111/dewb.12170]
59. Fu G, Ma L. Serious Violent Crimes and The Social Contradictions from the Perspective of Judicial Data. Journal of Law Application:Judicial Case (in Chinese) 2017;10:81-88.
60. Mei S, Li Z, Zhang X. Qualitative Data Analysis of 228 Cases of Workplace Violence on Medical Staffs Based on Internet Media. Chin Health Serv Manag (in Chinese) 2019;36(6):439-442.

Abbreviations

ITSA: interrupted time series analysis

Edited by G Eysenbach; submitted 02.05.20; peer-reviewed by H Chen, Y Liao; comments to author 16.06.20; revised version received 04.10.20; accepted 17.01.21; published 16.02.21.

Please cite as:
Yang Q, Tai-Seale M, Liu S, Shen Y, Zhang X, Xiao X, Zhang K
Measuring Public Reaction to Violence Against Doctors in China: Interrupted Time Series Analysis of Media Reports
J Med Internet Res 2021;23(2):e19651
URL: http://www.jmir.org/2021/2/e19651/
doi:10.2196/19651
PMID:33591282

Review

Effect of Telemetric Interventions on Glycated Hemoglobin A1c and Management of Type 2 Diabetes Mellitus: Systematic Meta-Review

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

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

Corresponding Author:
Claudia Eberle, MD, Prof Dr
Medicine with Specialization in Internal Medicine and General Medicine
Hochschule Fulda–University of Applied Sciences
Leipziger Strasse 123
Fulda, 36037
Germany
Phone: 49 661 9640 ext 6328
Fax: 49 661 9640 649
Email: claudia.eberle@hs-fulda.de

Abstract

Background: Diabetes mellitus is a chronic burden, with a prevalence that is increasing worldwide. Telemetric interventions have attracted great interest and may provide effective new therapeutic approaches for improving type 2 diabetes mellitus (T2DM) care.

Objective: The objective of this study was to analyze the clinical effectiveness of telemetric interventions on glycated hemoglobin A_{1c} (HbA_{1c}) specifically and T2DM management generally in a systematic meta-review.

Methods: A systematic literature search was performed in PubMed, CINAHL, Cochrane Library, Web of Science Core Collection, and EMBASE databases from January 2008 to April 2020. Studies that addressed HbA_{1c}, blood pressure, fasting blood glucose, BMI, diabetes-related and health-related quality of life, cost-effectiveness, time savings, and the clinical effectiveness of telemetric interventions were analyzed. In total, 73 randomized controlled trials (RCTs), 10 systematic reviews/meta-analyses, 9 qualitative studies, 2 cohort studies, 2 nonrandomized controlled studies, 2 observational studies, and 1 noncontrolled intervention study were analyzed.

Results: Overall, 1647 citations were identified. After careful screening, 99 studies (n=15,939 patients; n=82,436 patient cases) were selected by two independent reviewers for inclusion in the review. Telemetric interventions were categorized according to communication channels to health care providers: (1) “real-time video” interventions, (2) “real-time audio” interventions, (3) “asynchronous” interventions, and (4) “combined” interventions. To analyze changes in HbA_{1c}, suitable RCTs were pooled and the average was determined. An HbA_{1c} decrease of –1.15% (95% CI –1.84% to –0.45%), yielding an HbA_{1c} value of 6.95% (SD 0.495), was shown in studies using 6-month “real-time video” interventions.

Conclusions: Telemetric interventions clearly improve HbA_{1c} values in both the short term and the long term and contribute to the effective management of T2DM. More studies need to be done in greater detail.

(J Med Internet Res 2021;23(2):e23252) doi:10.2196/23252

KEYWORDS

telemedicine; telemetry; diabetes

Introduction

Diabetes mellitus is a chronic burden, with a prevalence that is increasing worldwide [1]. In 2019, approximately 463 million adults were diagnosed with diabetes [1]. By 2045, the International Diabetes Federation (IDF) projects an increase of 51% up to approximately 700 million people diagnosed with diabetes [1]. The IDF also estimates that one-half of individuals living with diabetes are undiagnosed [1]. According to the American Diabetes Association, type 2 diabetes mellitus

(T2DM) is the most prevalent type of diabetes and represents approximately 90% to 95% of all diabetes cases [2]. Common risk factors that appear to lead to T2DM are increasing age, increasing BMI, and lack of physical activity [2]. From a pathophysiological perspective, T2DM emerges mainly because of the progressive loss of beta-cell insulin secretion due to insulin resistance. Typically, however, relative insulin deficiency, as well as central and peripheral insulin resistance, arises [2].

T2DM is closely associated with diabetic microvascular complications—such as nephropathy, retinopathy, and neuropathy [3]—and macrovascular complications—such as coronary heart disease, stroke, and peripheral artery disease [4], as well as other comorbidities and general complications. In addition, cardiovascular disease is the main cause of death in patients with T2DM [4].

Therefore, optimal glycemic management is crucial [3]. Recent studies have reported positive effects of telemetric interventions on diabetes management [5,6]. Telemetry, defined as "a mode of delivering healthcare services through the use of telecommunications technologies, including but not limited to asynchronous and synchronous technology, and remote patient monitoring technology, by a healthcare practitioner to a patient or a practitioner at a different physical location than the healthcare practitioner" [7], may be a promising approach to improve the clinical effectiveness of T2DM management. This digital field of application is constantly evolving and expanding [8]. Telematics, the science of telecommunication and informatics, developed in the 1970s, and telemedicine emerged as a part of telematics in the 1970s and 1980s [8]. For a long time, the physical distance between the user groups was the dominant characteristic of telemedicine. The emergence of the internet in the 1990s opened up new communication channels. As a result, the focus was no longer on distance but on the fundamental application of technologies to overcome distance [8]. Electronic health (eHealth), characterized as health management based on electronic systems and communication, emerged from this idea [8]. The new concept of digital health combines the digital and genomic-proteomic revolutions with health care and everyday life [8].

In this systematic meta-review [9], we focused on telemetric communication pathways between health care professionals and patients. We aimed to update the evidence for and clinical effectiveness of telemetric approaches in the context of T2DM management considering different study designs such as randomized controlled trials (RCTs), clinical trials (CTs), systematic reviews (SRs), and meta-analyses (MAs). Furthermore, we focused on main clinical outcomes, such as glycated hemoglobin A_{1c} (HbA_{1c}), blood pressure (BP), fasting blood glucose (FBG), BMI, diabetes-related quality of life (DRQoL), and health-related quality of life (HRQoL), as well as the cost-effectiveness, time savings, and clinical effectiveness of telemetric interventions in general. HbA_{1c} is one of the major clinical parameters in T2DM and therefore our main focus.

To our knowledge, this study is the first and only systematic meta-review of telemetric interventions in T2DM management with respect to the following special features: we developed and applied a unique classification system for analyzing telemetric interventions and provide detailed insights by including several study designs and a wide range of clinical outcomes.

Research Design and Methods

Search Strategy

A systematic search was conducted targeting the period between January 2008 and April 2020. No protocol has been published. Keywords (diabetes mellitus, telemetry, telemonitoring, and telemedicine) were selected from the MEDLINE Medical Subject Headings and EMBASE Subject Headings databases and searched in titles/abstracts (Multimedia Appendix 1). In general, the steps were as follows: (1) search in five relevant databases, (2) eliminate duplicates, (3) screen titles and abstracts, (4) assess peer-reviewed publications for eligibility, (5) perform additive research via reference lists, (6) select T2DM studies, (7) extract relevant data, and (8) classify the publications.

Study Selection

Publications addressing telemetric interventions targeting T2DM management were included.

Telemetry was defined as "a mode of delivering healthcare services through the use of telecommunications technologies, including but not limited to asynchronous and synchronous technology, and remote patient monitoring technology, by a healthcare practitioner to a patient or a practitioner at a different physical location than the healthcare practitioner" [7]. We included video consultations, telephone counselling, asynchronous communication by email, SMS text messaging, internet/web-based platforms, and mixed forms.

Studies were screened and selected by two independent reviewers. Disagreements were resolved by a consensus-based discussion. We selected studies that met the following inclusion criteria: (1) peer-reviewed articles and studies; (2) written in English or German; (3) study design was an SR, MA, CT, or RCT; and (4) included interventions that involved direct interaction between patients and health care professionals through feedback and data transmission. We also considered quantitative and qualitative studies.

Smartphone/mobile app–based interventions were excluded and analyzed separately in another publication. We also rejected publications that observed mixed populations (eg, pooled patients with T1DM and T2DM), provided pooled data with other digital applications, addressed prevention or diagnosis, or focused on the presentation of technologies. Multimedia Appendix 2 shows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.

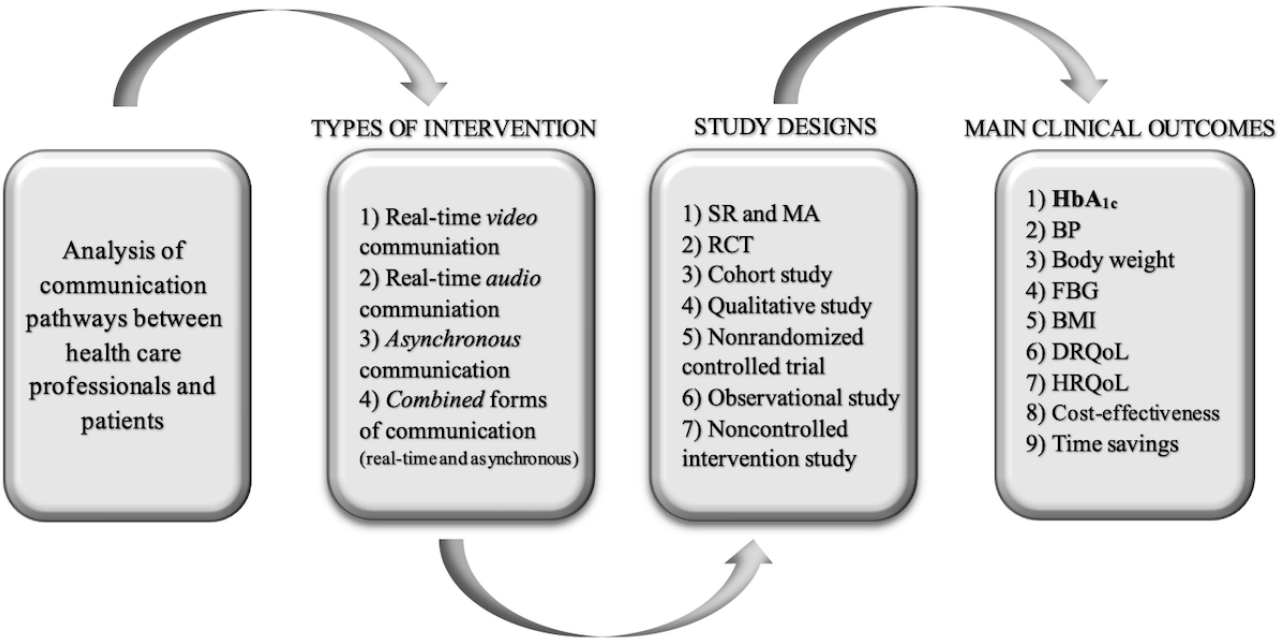
Data Extraction

Year of publication, location of the study, duration of the intervention, study design, sample sizes, intervention and control groups used, frequency of contact, feedback methods, outcomes, effects, statistical significance, and conclusions were extracted from each publication (step 7).

Study Classification and Analysis

For analysis, the interventions were classified a priori (step 8) based on the technologies used, study design, and outcomes (Figure 1).

Figure 1. Study classification procedure. BP: blood pressure; DRQoL: disease-related quality of life; FBG: fasting blood glucose; HbA_{1c}: glycated hemoglobin A1c; HRQoL: health-related quality of life; MA: meta-analysis; RCT: randomized controlled trial; SR: systematic review.



- "Real-time video" interventions (12 intervention studies): synchronous, face-to-face communication by videoconferencing and video consulting.
- "Real-time audio" interventions (17 intervention studies): synchronous communication by telephone calls (telephone coaching and counselling).
- "Asynchronous" interventions (28 intervention studies): asynchronous communication by email, SMS text messaging, internet/web-based platforms, server, home gateway, and post.
- "Combined" interventions (33 intervention studies): interventions involving real-time (ie, synchronous) and asynchronous communication, with a subgroup of "video clips" (interventions providing educational videos).

We conducted a small subgroup MA to assess whether the impact of the four intervention types, as well as the short- and long-term effects on the management of HbA_{1c} concentrations, differed. To determine the change in HbA_{1c}, we pooled appropriate RCTs and calculated the differences in means and 95% CIs for the intervention and control groups at the study end points. RCTs in which the changes from baseline to the end of the study were reported as a percentage were included. Studies in which the control group received telemetric support were excluded. Mean deviations and SDs were extracted unchanged.

In addition, the publication bias was assessed visually as a funnel plot using HbA_{1c} values based on the RCTs and the mean differences (MDs) from our subgroup MA.

We also pooled the number of patients, specifically the number of unique patients as well as the number of patient cases related to the outcomes. In the former scenario, each patient occurred only once, addressing the number of individual patients (without SRs and MAs), and in the latter scenario, with a focus on specific outcomes, patient cases were analyzed based on the respective outcomes and thus may have been included several times (including SRs and MAs).

Results

Description of Studies

Our search strategy identified 1647 citations. After removing duplicates, 1116 studies were screened and 875 ineligible papers excluded. After assessing 241 studies with full text, 72 inappropriate studies were rejected. As an interim result, 189 studies were identified, of which 23 focused on type 1 diabetes mellitus, 99 focused on T2DM, 11 focused on gestational diabetes, and 51 focused on mixed populations. In this systematic meta-review, we included 99 suitable T2DM publications, analyzing 15,939 patients and 82,439 patient cases. A list of the included studies is provided in Multimedia Appendix 3.

Baseline characteristics of the studies are summarized in Table 1. Of the 99 studies, 10 were SRs and MAs, 73 were RCTs, 9 were qualitative examinations, 2 were cohort studies, 2 were non-RCTs, 2 were observational studies, and 1 was a noncontrolled intervention study. When classifying the studies according to location and type of intervention, SRs and MAs were excluded due to their heterogeneity, and thus 89 studies were taken into account. Of these 89 studies, 35 were done in the United States, 21 in Asia, 20 in Europe, 6 in Australia, 3 in Canada, 2 in Brazil, and 2 in Turkey.

In total, 12 "real-time video," 17 "real-time audio," 28 "asynchronous," 33 "combined," and 3 "video clip"

interventions were classified. One study matched the classification criteria for two categories [10]. A detailed summary of all studies is shown in Multimedia Appendix 4.

A descriptive examination of the funnel plot created using HbA_{1c} values indicated a mild form of asymmetry (Multimedia Appendix 8).

Table 1. Baseline characteristics of reviewed studies.

Studies	n (%)
All studies (N=99)	
Study design	
SRs[a] and MAs[b]	10 (10)
Randomized controlled trials (total)	73 (74)
Pilot studies	3 (3)
Cohort studies	2 (2)
Qualitative studies	9 (9)
Nonrandomized controlled trials	2 (2)
Observational studies	2 (2)
Noncontrolled intervention studies	1 (1)
Years	
2008-2011	21 (21)
2012-2014	26 (26)
2015-2017	32 (32)
2018-2020	19 (19)
All studies, excluding SRs and MAs (n=89)	
Location	
United States	35 (39)
Canada	3 (3)
Brazil	2 (2)
Europe	20 (23)
Asia	21 (24)
Australia	6 (7)
Turkey	2 (2)
Intervention	
Real-time video	12 (14)
Real-time audio	17[c] (19)
Asynchronous	28[c] (32)
Combined forms (total)	33 (37)
"Video clips" subgroup	3 (3)

[a]SRs: systematic reviews.

[b]MAs: meta-analyses.

[c]One study matched the criteria for two categories.

Impact on Main Outcomes

An overview of significant and not significant intervention effects on HbA_{1c}, BP, FBG, BMI, DRQoL, HRQoL, cost-effectiveness, time savings, and clinical effectiveness is displayed in Multimedia Appendix 5. Multimedia Appendix 6 shows the significant effects on the main outcomes. Briefly, 85% (84/99) of the intervention studies found explicit beneficial effects due to telemetric interventions, depending on the outcomes studied (see Multimedia Appendix 4).

SRs and MAs (n=10)

HbA_{1c} (n=8)

All SRs and MAs reported clear decreases in HbA_{1c} values ($P<.05$) by implementing telemetric interventions [5,11-17]. MAs (5/5, 100%) indicated that telemetry was significantly associated with an obvious improvement between –0.37% and –0.55% in HbA_{1c} values compared with the usual care ($P<.001$ [15], $P<.001$ [13], $P<.001$ [12], $P<.001$ [5], and $P<.05$ [10]).

BMI (n=1)

According to Kim et al [5], telemonitoring was associated with a significantly reduced BMI (weighted MD=–0.25 kg/m², 95% CI –0.49 to –0.01, $I^2=16.7\%$) compared with usual care.

Cost-Effectiveness (n=1)

Due to the heterogeneous data situation, Zhai et al [13] could not draw any conclusions regarding cost-effectiveness.

"Real-Time Video" Interventions (n=12)

HbA_{1c} (n=9)

Overall, 89% (8/9) of the studies reported a clear reduction in HbA_{1c} values. More specifically, five RCTs (5/9, 56%) indicated significant positive effects ($P=.022$ [18], $P=.004$ [19], $P=.023$ [20], $P<.05$ [21], and $P=.013$ [22]). For example, HbA_{1c} values declined significantly in an intervention group with weekly video conferences compared with a control group (0.49% versus 0.17%; $P=.013$) [22].

FBG (n=3)

Overall, definite improvements regarding FBG were documented. Tavsanli et al [22] (weekly video conferences) and Rasmussen et al [20] (average 4.1 video consultations in 6 months) reported clearly lower FBG levels in the intervention group compared with control groups ($P>.05$ [22] and $P<.015$ [20]), whereas Hansen et al [18] provided monthly video conferences additional to usual care and reported no substantial changes in FBG levels in relation to the study (significance not reported).

BP (n=3)

In general, most RCTs observed no essential changes in systolic and diastolic BP measurements (intergroup $P>.05$ [20] and significance not reported [18]). However, a clear improvement in BP was seen in a 12-month videoconferencing intervention, as reported by Davis et al [19], although the effect was not significant compared with a control group (intervention systolic BP 130.8 mmHg, SD 3.6 mmHg, versus 127.6 mmHg, SD 4.0 mmHg, $P=.76$; diastolic BP 72.7 mmHg, SD 2.1 mmHg, versus 70.2 mmHg, SD 2.2 mmHg, $P=.64$).

Body Weight (n=1)

Rasmussen et al [20] showed a significantly higher weight loss with in-person clinic visits (–1.7 kg) compared with video consultations (–0.6 kg; $P=.023$).

BMI (n=2)

Hansen et al [18] found no obvious changes in terms of BMI, whereas Davis et al [19] indicated substantial improvements compared to usual care, although the finding was not significant (30.6 kg/m², SD 1.4 kg/m² versus 35.8 kg/m², SD 1.4 kg/m²; $P=.73$).

HRQoL (n=1)

Hansen et al [18] noted that significant changes in mental or physical health rankings were not detected.

Time Savings (n=1)

Gordon et al [23] revealed shorter travel times and less time in waiting rooms according to interviews with participants, although no statistical measurements were performed.

Enablers and Barriers (n=2)

Carlisle and Warren [24] suggested that consumer-friendly technologies and the integration of telemetry into everyday lives are important for the successful implementation of telemetry interventions.

"Real-Time Audio" Interventions (n=17)

HbA_{1c} (n=9)

In summary, all studies showed precise improvements in HbA_{1c} levels with audio interventions in real time. Odnoletkova et al [25] and Walker et al [26] reported significant improvements in their intervention groups compared with matched control groups (intervention group: –0.2%, 95% CI –0.3 to –0.1, $P=.003$ [25]; and intervention group versus control group: –0.23% versus 0.13%, $P=.04$ [26]). Sarayani et al [27], Trief et al [28], Maslakpak et al [29], Blackberry et al [30], Benson et al [31], and Vasconcelos et al [32] displayed clear, but not significant, improvements in HbA_{1c} levels compared with control groups ($P>.05$). Notably, some control groups [27-29,31] received some forms of telemetric or even educational support. Interestingly, Walker et al [26] found that patients who completed at least six phone calls with a health educator over a 12-month period had significant reductions in their HbA_{1c} concentrations ($P<.05$).

The RCT of McMahon et al [10] was classified in two categories ("real-time audio" and "asynchronous" interventions) because it involved the comparison of a telephone-based intervention with two "asynchronous" interventions. HbA_{1c} decreased at a rate of 0.32% every 3 months for the online arm, 0.36% for the telephone arm, and 0.41% for the web training arm (all $P<.001$).

FBG (n=2)

The "real-time audio" intervention studies showed obvious improvements in FBG. In the study by Varney et al [33], FBG levels clearly improved in subjects in the intervention group (8.9 mmol/L, 95% CI 8.0 to 9.7, to 8.5 mmol/L, 95% CI 7.7 to 9.4) compared with those in the control group ($P=.02$), but not in a long-term way, while Maslakpak et al [29] outlined distinct, but not significant, differences between telephone and control groups ($P=.766$).

BP (n=3)

In general, all "real-time audio" intervention studies reported clear improvements in BP. Trief et al [28] showed a greater improvement in systolic BP in the "individual calls" group than

in the "diabetes education" group at 8 months (P=.021). Vasconcelos et al [32] reported improvements in systolic BP (130.25 mmHg to 125.87 mmHg; P=.171) and diastolic BP (72.12 mmHg to 71.12 mmHg; P=.640). In addition, Varney et al [33] indicated significant improvements in diastolic BP within the telephone group (80 mmHg, 95% CI 76 to 84, to 74 mmHg, 95% CI 71 to 77), but these were not sustained.

Body Weight (n=1)

According to Odnoletkova et al [25], the difference between the groups in favor of telecoaching was a change in body weight of –1.1 kg (P=.004).

BMI (n=3)

In general, all trials noted slight improvements in BMI. Odnoletkova et al [25] and Trief et al [28] reported significant improvements between groups (P=.003 [25] and P=.021 [28]), whereas Vasconcelos et al [32] indicated a slight decrease (29.99 kg/m² to 29.96 kg/m²) that was not significant (P=.764).

Cost-Effectiveness (n=2)

"Real-time audio" interventions appear to be moderate in terms of cost-effectiveness (no statistical significances reported). Schechter et al [34] concluded that the costs were moderate relative to the benefits, whereas Varney et al [35] revealed that the cost of a 10-year intervention was covered by the financial savings, with a tendency for health profits.

"Asynchronous" Interventions (n=28)

HbA_{1c} (n=24)

Overall, the majority of the studies (all RCTs; 23/34, 96%) reported apparent improvements. Eleven RCTs reported significant improvements in HbA_{1c} in the intervention groups compared with the control groups (P<.05) [36-46], whereas 3 RCTs showed significant beneficial effects within their intervention groups (P<.05) [47-49]. In addition, 5 RCTs found improvements in the intervention groups compared with matched controls, but the results were not statistically significant (P>.05) [50-54]. Ramadas et al [55], Tildesley et al [38], and Cho et al [56] mentioned significant improvements within their intervention groups (P=.004 [55]; real-time continuous glucose monitoring, P<.001, versus internet blood glucose monitoring system, P<.05 [38]; and P<.01 [56]), but the differences between groups were not significant (P>.05).

FBG (n=2)

The studies showed significant improvements in FBG levels in the intervention groups compared with control groups (8.9 mmol/L, SD 3.9 mmol/L, versus 7.9 mmol/L, SD 2.5 mmol/L, P=.015 [55]; and P=.005 [46]).

BP (n=3)

In summary, the publications reported apparent improvements in BP. Wild et al [39] found significant improvements in systolic BP (P=.017) and diastolic BP (P=.006) in the intervention group compared with the control group. Wakefield et al [40] also found a significant decrease in systolic BP in the high-intensity arm of the intervention (home telehealth device with algorithm; P=.01). Fang and Deng [44] found improvements in systolic BP (P=.069) and diastolic BP (P=.693) in the treatment group, but they were not significant.

Body Weight (n=3) and BMI (n=2)

In general, the studies revealed clear beneficial effects of "asynchronous" interventions on both body weight and BMI. For example, Luley et al [41] showed large significant improvements in body weight (–11.8 kg, SD 8.0 kg; both inter- and intragroup comparisons with P=.000) and BMI (–4.1 kg/m²; both intergroup and intragroup comparisons with P=.00).

HRQoL (n=1)

No clinically important improvements in HRQoL were seen according to Dario et al [51].

Cost-Effectiveness (n=1)

A weight-loss telemonitoring intervention from Luley et al 2011 [41] showed an effective decline in medication costs of €83 (US $101) per patient in 6 months.

Time Savings (n=1)

Cho et al [57] showed a significant time savings for physicians of approximately 55% focusing on patients with HbA_{1c} levels greater than 6.5% (P<.05).

"Combined" Interventions (n=33)

HbA_{1c} (n=24)

In general, most publications (21/24, 88%) reported clear significant improvements in HbA_{1c} (P<.05). Three of these studies were RCTs that achieved significant improvements within their intervention groups (P<.001 [58]; P=.27 [59]; and P value not reported [60]) but not significant differences between the intervention and control groups (P>.05).

FBG (n=6)

Overall, the studies showed mostly positive effects of "combined" interventions on FBG [58,61-64]. For example, Zhou et al [64] and Jeong et al [58] found significant reductions in FBG levels compared with the control groups (8.73 mmol/L to 7.06 mmol/L, P<.001 [64]; and –12.28 mg/dL, SD 41.20 mg/dL, P=.027 (telemedicine group [58]).

BP (n=13)

Approximately 85% (11/13) of the "combined" intervention studies reported beneficial effects on BP [59-61,63-70]. Kempf et al [70] and Crowley et al [65] found significant improvements compared with control groups (systolic BP, P<.01 [70]; and systolic BP, P=.035, and diastolic BP, P=.013 [65]). However, some RCTs noted improvements in their intervention groups (ie, P<.05) but no significant differences between the groups (P>.05) [60,66-69,71]. Additionally, Kesavadev et al (cohort study with 1000 participants [63]) and Dienstl et al (observational study [61]) indicated similar significant improvements in systolic and diastolic BP (P<.01 [63] and P<.001 [61]).

Body Weight (n=7)

Most "combination" intervention studies (5/7, 71%) found clear improvements in body weight [60,61,69,70,72]. For example,

Kempf et al [70] reported a significant reduction of 6.2 (SD 4.6) kg in the intervention group compared with the control group (–1.0 kg, SD 3.4 kg; *P*<.01).

BMI (n=9)

The majority of publications (7/9, 78%) showed an apparent reduction of BMI. Significant improvements were outlined by 6 studies (intragroup *P*=.047 [59], intragroup *P*<.01 [63], intragroup *P*<.001 [61], intergroup *P*=.036 [73], intergroup *P*<.01 [70], and intergroup *P*<.05 [74]). For example, Kesavadev et al [63] (n=1000 patients) showed a significant reduction of 0.3 kg/m² (*P*<.01) and Kempf et al [70] reported –2.1 (SD 1.5) kg/m² in the intervention group versus –0.3 (SD 1.1) kg/m² in the control group (*P*<.01).

DRQoL (n=3) and HRQoL (n=1)

All studies found clear improvements in DRQoL and HRQoL. Kempf et al [70] and Nicolucci et al [69] showed significant intergroup improvements in HRQoL (*P*<.01 and *P*<.03, respectively). Jha et al [62] and Dienstl et al [61] (observational studies) reported significant beneficial effects with regard to DRQoL (*P*=.015 and *P*<.001, respectively).

Cost-Effectiveness (n=2)

Warren et al [75] and Kesavadev et al [63] reported that "combined" interventions are cost-effective. The total costs for the internet-based treatment group were lower than those for the control group (mean US $3781 versus US $4662; *P*<.001 [75]). According to Kesavadev et al [63], the extra cost was US $9.66/month (significance not reported), but money and time saved in physical visits made up for the extra costs.

Time Savings (n=1)

Hsu et al [72] reported great time savings with a cloud-based diabetes management program compared with standard face-to-face care (22.5-minute versus 68.8-minute visit time; significance not reported).

"Combined" Interventions—"Video Clips" Subgroup (n=3)

HbA_{1c} (n=3)

All studies reported significant reductions in HbA_{1c} compared with control subjects (*P*<.001 [76], *P*=.005 [77], and *P*=.013 [78]).

BP and Body Weight (n=1)

Tang et al [76] detected improvements in BP (systolic BP, *P*=.306, and diastolic BP, *P*=.374) but no effects on body weight (*P*=.232).

Short- and Long-Term Effects on HbA_{1c} Values (n=41)

Short- and long-term effects based on the comparison of HbA_{1c} values between the intervention and control groups at the study end points were investigated. Patients' changes in HbA_{1c} from baseline to the end of the study of 41 RCTs are presented in Multimedia Appendix 7.

"Real-Time Video" Interventions

A small MA showed that, compared with the control group, 6-month interventions (n=2) were associated with a greater effect size (MD=–1.15%, 95% CI –1.84 to –0.45) than 12-month interventions (n=2) (MD=–0.6%, 95% CI –0.99 to –0.21).

"Real-Time Audio" Interventions

The subgroup analysis revealed an effect size, compared to usual care, of MD=–0.37% (95% CI –0.79 to 0.05) for 6-month interventions (n=3) compared with –0.5% in the 3-month intervention [29] and –0.06% in the 18-month intervention [30].

"Asynchronous" Interventions

The greatest effect was seen in 12-month interventions (n=2) (MD=–0.77%, 95% CI –2.25 to 0.72), followed by 6-month interventions (n=8) (MD=–0.57%, 95% CI –0.75 to –0.39), and 3-month interventions (n=3) (MD=–0.38%, 95% CI –0.54 to –0.22).

"Combined" Interventions

The 3-month interventions (n=5) had the greatest effect (MD=–0.65%, 95% CI –0.98 to –0.31), whereas 6-month interventions (n=7) had a slightly smaller effect (MD=–0.50%, 95% CI –0.71 to –0.30). In comparison, the effect of 12-month interventions (n=4) was even smaller (MD=–0.25%, 95% CI –0.73 to 0.24). The subgroup "video clip" interventions (n=2) showed a reduction of MD=–0.23 (95% CI –0.23 to –0.23).

Discussion

Principal Results

Telemetry is a viable alternative to usual care for patients with T2DM and can lead to improvements in a wide range of outcomes. The inclusion of evidence from different study designs, such as reviews and trials, in our review strengthens the conclusion that use of telemetric interventions can be feasible in a clinical setting. Other reviews have also recently presented an improvement of clinical outcomes through telemetry and especially a trend toward a reduction in HbA_{1c} levels [16,79]. Our results suggest that telemetry generated clinically meaningful reductions in HbA_{1c} levels. Telemetry has the advantage of helping people who are restricted due to geographic location or a lack of resources [14]. In the time of COVID-19 in particular, the advantages and potential of remote diabetes management becomes even more important.

Impact of Telemetric Interventions on HbA1c

In general, all types of telemetric interventions clearly improved HbA_{1c}. All SRs and MAs also clearly showed that telemetric interventions improve HbA_{1c} specifically, as well as the management of T2DM generally. Furthermore, "real-time video" interventions with a duration of 6 months were the most effective in reducing HbA_{1c}. These interventions showed clear improvements in HbA_{1c} levels in patients diagnosed with T2DM compared to usual care (MD=–1.15%, 95% CI –1.84 to –0.45). Overall, the effects in the subgroup analysis in terms of the improvement of HbA_{1c} values had MDs between –1.15% and –0.25%. These obvious decreases in HbA_{1c} may indicate a novel

and additional approach to diabetes care since these therapeutic effects could be accomplished by telemetric intervention alone. However, to optimize glucose homeostasis, individual telemetric approaches may be considered in terms of individual diabetes care as an addition to established therapeutic approaches [4].

Impact of Telemetric Interventions on Main Clinical Outcomes

Through the use of "combined" interventions, FBG levels improved effectively, which was shown by a moderate number of studies. With "asynchronous" and "real-time audio" interventions, few studies showed an improvement in FBG values. However, the data were inconsistent as to whether "real-time video" interventions reduce FBG effectively.

BP measurements decreased by applying "combined" interventions in a moderate number of the reviewed studies. "Asynchronous" and "real-time audio" interventions also improved BP, but there were comparatively few examinations. Moreover, the study situation for "real-time video" interventions was found to be rather inconsistent.

Body weight decreased in a moderate number of studies by using "combined" interventions effectively. "Real-time audio" interventions also clearly reduced body weight in a few investigations. However, the study situation for "real-time video" interventions was not consistent.

BMI decreased effectively in several studies by using "combined" interventions. The few studies available indicated that "asynchronous" and "real-time audio" interventions decreased BMI. In contrast, the study situation for "real time video" was inconsistent.

Only "combined" interventions showed effective improvements regarding DRQoL and HRQoL, but there were few studies that examined DRQoL and very few studies of HRQoL compared with the other clinical outcomes. "Real-time audio," "asynchronous," and "combined" interventions were potentially cost-effective, but there was only a small number of studies. In addition, "real-time video," "asynchronous," and "combined" interventions occasionally showed time savings, although again, few studies examined these outcomes.

Impact and Comparison of Different Telemetric Interventions

"Real-time video" interventions did improve HbA_{1c} clearly and effectively in short-term and long-term ways in a large number of studies. Weekly videoconferencing seems to be very effective in terms of reducing HbA_{1c}. Due to the heterogeneity of the studies, the results regarding FBG, BP, body weight, BMI, and QoL may be rather inconsistent. However, they all have in common that user-friendly technologies were considered in the development of the interventions and that telemetry was anchored in people's everyday lives, both of which are necessary for optimal results.

"Real-time audio" interventions proved to be effective in reducing HbA_{1c}, as demonstrated in numerous studies. Some studies indicated that there is also a clear beneficial impact of these interventions on FBG, BP, body weight, and BMI. Additionally, "real-time audio" interventions were shown to be cost-effective by the limited studies available.

Furthermore, a large number of studies pointed out that "asynchronous" interventions improved HbA_{1c} effectively. These interventions also improved FBG, BP, and BMI, and showed very positive results in terms of cost-effectiveness and time savings, but few studies using "asynchronous" interventions were available for review.

Most studies assessed "combined" interventions (real-time and asychronous communication). Numerous studies indicated that "combined" interventions improved HbA_{1c} values effectively. Furthermore, a moderate number of interventions had a favorable impact on FBG, BP, BMI, and body weight. In terms of DRQoL and HRQoL, there were few studies to examine these outcomes, but the available studies showed positive tendencies. Additionally, cost-effectiveness and time savings of telemetric interventions showed a positive trend, but sufficient data were lacking.

From our point of view, telemetric T2DM management enhances patient compliance, enables intensive monitoring, and empowers patients to deal with and understand their disease. For a successful implementation of telemetric approaches, it is also essential that the technology is user-friendly, that telemetric T2DM management can be easily integrated into everyday life, and that it is tailored to the patient and his or her life circumstances [24].

Furthermore, we would like to point out that telemetric interventions differ not only in terms of their technologies but also in terms of their contextual focus (eg, nutrition, exercise, etc) and that this aspect should be taken into account when interpreting the results.

Study Limitations

Although the exclusion criteria were observed, the included studies displayed a wide variation in terms of study design, technical and interventional approaches, duration, and frequency of contact with health care providers (in both the intervention and the control groups), as well as sample size and statistical evaluations used. Due to this heterogeneity, as well as to the small size of our MA, there may be potential for bias. For the same reasons, methodological quality and statistical evaluations could not be carried out. Some studies achieved improvements that were significant within the intervention groups but not between the groups, and methodological weaknesses may have been responsible for that.

Comparison With Prior Work

Other research groups have displayed similar results. Numerous other SRs and MAs, which were included in this review, reported significant decreases in HbA_{1c} values ($P<.05$) from the implementation of telemetric interventions [5,11-17]. Su et al [79] examined 55 RCTs and concluded that telemedicine effectively improved clinical outcomes as well as T2DM management compared with usual care. Lee et al [16], who included 4 SRs reporting on 29 studies, concluded that telemetry was a very effective therapeutic approach in terms of decreasing HbA_{1c}. According to a review and network MA by Lee et al

[14], over a 6-month follow-up, telemedicine reduced HbA_{1c} by a mean of 0.43% (95% CI −0.64% to −0.21%). The authors concluded that all telemedical strategies, with the exceptions of telecase management and telementoring, were effective in reducing HbA_{1c} in a clinically meaningful way. Furthermore, Mushcab et al [17] showed that telemonitoring effectively improved HbA_{1c} levels and quality of life. They also observed a high acceptance of web-based systems.

Our research builds on these previous findings, incorporating a large number of studies (n=99), patients (n=15,939), and patient cases (n=82,436) and considering a range of main clinical outcomes in terms of T2DM management. Interestingly, there may be differences in telemetric approaches in terms of T2DM versus type 1 diabetes mellitus management [79], but these still need to be analyzed in more detail.

Conclusions

To our knowledge, this is the first systematic meta-review analyzing telemetric approaches in T2DM management, including a wide range of important clinical outcomes and technologies.

Viewed together, telemetric interventions clearly improve HbA_{1c} values in the short term and long term specifically and T2DM care generally. Moreover, “real-time video” interventions with a duration of 6 months showed the greatest effect in terms of improving HbA_{1c} values in a sustained way. “Combined” interventions (real-time and asynchronous communication) appeared to be most effective in improving FBG, BP, body weight, BMI, and quality of life.

In conclusion, telemetric interventions clearly improve HbA_{1c} and T2DM management effectively. More studies need to be done, especially with a focus on main clinical outcomes.

Acknowledgments

This research work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), project number EB 440/4-1. Therefore, the authors would like to thank the DFG for the strong support of this research work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search strings.
[PDF File (Adobe PDF File), 417 KB - jmir_v23i2e23252_app1.pdf]

Multimedia Appendix 2
Systematic meta-review search and selection procedure based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.
[PDF File (Adobe PDF File), 525 KB - jmir_v23i2e23252_app2.pdf]

Multimedia Appendix 3
List of included studies.
[PDF File (Adobe PDF File), 362 KB - jmir_v23i2e23252_app3.pdf]

Multimedia Appendix 4
Summary of the studies included in this systematic meta-review.
[PDF File (Adobe PDF File), 872 KB - jmir_v23i2e23252_app4.pdf]

Multimedia Appendix 5
Impact on main clinical outcomes, significant and not significant effects.
[PDF File (Adobe PDF File), 529 KB - jmir_v23i2e23252_app5.pdf]

Multimedia Appendix 6
Significant effects on main clinical outcomes.
[PDF File (Adobe PDF File), 424 KB - jmir_v23i2e23252_app6.pdf]

Multimedia Appendix 7
Changes in glycated hemoglobin (HbA1c) values (%) in intervention and control groups from baseline to end of the study (n=41 randomized controlled trials).
[PDF File (Adobe PDF File), 504 KB - jmir_v23i2e23252_app7.pdf]

Multimedia Appendix 8
Funnel plot using glycated hemoglobin (HbA1c) based on the randomized controlled trials from the subgroup meta-analysis.
[DOCX File , 16 KB - jmir_v23i2e23252_app8.docx]

References

1. International Diabetes Foundation. IDF Diabetes Atlas 9th edition. 2019. URL: https://www.diabetesatlas.org/en/resources/ [accessed 2020-01-28]
2. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. Diabetes Care 2020 Jan;43(Suppl 1):S14-S31. [doi: 10.2337/dc20-S002] [Medline: 31862745]
3. American Diabetes Association. 11. Microvascular Complications and Foot Care: Standards of Medical Care in Diabetes-2020. Diabetes Care 2020 Jan;43(Suppl 1):S135-S151. [doi: 10.2337/dc20-S011] [Medline: 31862754]
4. Viigimaa M, Sachinidis A, Toumpourleka M, Koutsampasopoulos K, Alliksoo S, Titma T. Macrovascular Complications of Type 2 Diabetes Mellitus. Curr Vasc Pharmacol 2020;18(2):110-116. [doi: 10.2174/1570161117666190405165151] [Medline: 30961498]
5. Kim Y, Park J, Lee B, Jung C, Park D. Comparative effectiveness of telemonitoring versus usual care for type 2 diabetes: A systematic review and meta-analysis. J Telemed Telecare 2019 Dec;25(10):587-601. [doi: 10.1177/1357633X18782599] [Medline: 30012042]
6. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth Interventions to Support Self-Management of Long-Term Conditions: A Systematic Metareview of Diabetes, Heart Failure, Asthma, Chronic Obstructive Pulmonary Disease, and Cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
7. American Telemedicine Association. ATA's Standardized Telehealth Terminologyand Policy Language for States on Medical Practice. 2020. URL: https://www.americantelemed.org/wp-content/uploads/2020/10/ATA-_Medical-Practice-10-5-20.pdf [accessed 2021-01-28]
8. Meister S, Becker S, Leppert F, Drop L. Digital Health, Mobile Health und Co. - Wertschöpfung durch Digitalisierung und Datenverarbeitung. In: Pfannstiel MA, Da-Cruz P, Mehlich H, editors. Digitale Transformation von Dienstleistungen im Gesundheitswesen I. Wiesbaden: Springer Fachmedien Wiesbaden; 2016:185-212.
9. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
10. McMahon GT, Fonda SJ, Gomes HE, Alexis G, Conlin PR. A randomized comparison of online- and telephone-based care management with internet training alone in adult patients with poorly controlled type 2 diabetes. Diabetes Technol Ther 2012 Nov;14(11):1060-1067 [FREE Full text] [doi: 10.1089/dia.2012.0137] [Medline: 22953754]
11. Huang Z, Tao H, Meng Q, Jing L. Management of endocrine disease. Effects of telecare intervention on glycemic control in type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. Eur J Endocrinol 2015 Mar;172(3):R93-101 [FREE Full text] [doi: 10.1530/EJE-14-0441] [Medline: 25227131]
12. Cassimatis M, Kavanagh DJ. Effects of type 2 diabetes behavioural telehealth interventions on glycaemic control and adherence: a systematic review. J Telemed Telecare 2012 Dec;18(8):447-450. [doi: 10.1258/jtt.2012.GTH105] [Medline: 23209266]
13. Zhai Y, Zhu W, Cai Y, Sun D, Zhao J. Clinical- and cost-effectiveness of telemedicine in type 2 diabetes mellitus: a systematic review and meta-analysis. Medicine (Baltimore) 2014 Dec;93(28):e312 [FREE Full text] [doi: 10.1097/MD.0000000000000312] [Medline: 25526482]
14. Lee SWH, Chan CKY, Chua SS, Chaiyakunapruk N. Comparative effectiveness of telemedicine strategies on type 2 diabetes management: A systematic review and network meta-analysis. Sci Rep 2017 Oct 04;7(1):12680. [doi: 10.1038/s41598-017-12987-z] [Medline: 28978949]
15. Greenwood DA, Young HM, Quinn CC. Telehealth Remote Monitoring Systematic Review: Structured Self-monitoring of Blood Glucose and Impact on A1C. J Diabetes Sci Technol 2014 Feb 21;8(2):378-389. [doi: 10.1177/1932296813519311] [Medline: 24876591]
16. Lee PA, Greenfield G, Pappas Y. The impact of telehealth remote patient monitoring on glycemic control in type 2 diabetes: a systematic review and meta-analysis of systematic reviews of randomised controlled trials. BMC Health Serv Res 2018 Jun 26;18(1):495 [FREE Full text] [doi: 10.1186/s12913-018-3274-8] [Medline: 29940936]
17. Mushcab H, Kernohan WG, Wallace J, Martin S. Web-Based Remote Monitoring Systems for Self-Managing Type 2 Diabetes: A Systematic Review. Diabetes Technol Ther 2015 Jul;17(7):498-509. [doi: 10.1089/dia.2014.0296] [Medline: 25830528]
18. Hansen CR, Perrild H, Koefoed BG, Zander M. Video consultations as add-on to standard care among patients with type 2 diabetes not responding to standard regimens: a randomized controlled trial. Eur J Endocrinol 2017 Jun;176(6):727-736. [doi: 10.1530/EJE-16-0811] [Medline: 28325823]

19. Davis RM, Hitch AD, Salaam MM, Herman WH, Zimmer-Galler IE, Mayer-Davis EJ. TeleHealth improves diabetes self-management in an underserved community: diabetes TeleCare. Diabetes Care 2010 Aug;33(8):1712-1717 [FREE Full text] [doi: 10.2337/dc09-1919] [Medline: 20484125]
20. Rasmussen OW, Lauszus FF, Loekke M. Telemedicine compared with standard care in type 2 diabetes mellitus: A randomized trial in an outpatient clinic. J Telemed Telecare 2016 Sep;22(6):363-368. [doi: 10.1177/1357633X15608984] [Medline: 26468213]
21. Toledo FGS, Ruppert K, Huber KA, Siminerio LM. Efficacy of the Telemedicine for Reach, Education, Access, and Treatment (TREAT) model for diabetes care. Diabetes Care 2014 Aug;37(8):e179-e180. [doi: 10.2337/dc13-1909] [Medline: 25061149]
22. Tavsanly NG, Karadakova A, Saygili F. The use of videophone technology (telenursing) in the glycaemic control of diabetic patients: a randomized controlled trial. J Diab Res Clin Met 2013;2(1):1. [doi: 10.7243/2050-0866-2-1]
23. Gordon HS, Solanki P, Bokhour BG, Gopal RK. "I'm Not Feeling Like I'm Part of the Conversation" Patients' Perspectives on Communicating in Clinical Video Telehealth Visits. J Gen Intern Med 2020 Jun;35(6):1751-1758. [doi: 10.1007/s11606-020-05673-w] [Medline: 32016705]
24. Carlisle K, Warren R. A qualitative case study of telehealth for in-home monitoring to support the management of type 2 diabetes. J Telemed Telecare 2013 Oct;19(7):372-375. [doi: 10.1177/1357633X13506512] [Medline: 24218347]
25. Odnoletkova I, Goderis G, Nobels F, Fieuws S, Aertgeerts B, Annemans L, et al. Optimizing diabetes control in people with Type 2 diabetes through nurse-led telecoaching. Diabet Med 2016 Jun;33(6):777-785. [doi: 10.1111/dme.13092] [Medline: 26872105]
26. Walker EA, Shmukler C, Ullman R, Blanco E, Scollan-Koliopoulus M, Cohen HW. Results of a successful telephonic intervention to improve diabetes control in urban adults: a randomized trial. Diabetes Care 2011 Jan;34(1):2-7 [FREE Full text] [doi: 10.2337/dc10-1005] [Medline: 21193619]
27. Sarayani A, Mashayekhi M, Nosrati M, Jahangard-Rafsanjani Z, Javadi M, Saadat N, et al. Efficacy of a telephone-based intervention among patients with type-2 diabetes; a randomized controlled trial in pharmacy practice. Int J Clin Pharm 2018 Apr;40(2):345-353. [doi: 10.1007/s11096-018-0593-0] [Medline: 29435911]
28. Trief PM, Fisher L, Sandberg J, Cibula DA, Dimmock J, Hessler DM, et al. Health and Psychosocial Outcomes of a Telephonic Couples Behavior Change Intervention in Patients With Poorly Controlled Type 2 Diabetes: A Randomized Clinical Trial. Diabetes Care 2016 Dec;39(12):2165-2173 [FREE Full text] [doi: 10.2337/dc16-0035] [Medline: 27456837]
29. Maslakpak M, Razmara S, Niazkhani Z. Effects of Face-to-Face and Telephone-Based Family-Oriented Education on Self-Care Behavior and Patient Outcomes in Type 2 Diabetes: A Randomized Controlled Trial. J Diabetes Res 2017;2017:8404328.
30. Blackberry ID, Furler JS, Best JD, Chondros P, Vale M, Walker C, et al. Effectiveness of general practice based, practice nurse led telephone coaching on glycaemic control of type 2 diabetes: the Patient Engagement And Coaching for Health (PEACH) pragmatic cluster randomised controlled trial. BMJ 2013 Sep 18;347:f5272. [doi: 10.1136/bmj.f5272]
31. Benson GA, Sidebottom A, Hayes J, Miedema MD, Boucher J, Vacquier M, et al. Impact of ENHANCED (diEtitiaNs Helping pAtieNts CarE for Diabetes) Telemedicine Randomized Controlled Trial on Diabetes Optimal Care Outcomes in Patients with Type 2 Diabetes. J Acad Nutr Diet 2019 Apr;119(4):585-598. [doi: 10.1016/j.jand.2018.11.013] [Medline: 30711463]
32. de Vasconcelos HCA, Lira Neto JCG, de Araújo MFM, Carvalho GCN, de Souza Teixeira CR, de Freitas RWJF, et al. Telecoaching programme for type 2 diabetes control: a randomised clinical trial. Br J Nurs 2018 Oct 18;27(19):1115-1120. [doi: 10.12968/bjon.2018.27.19.1115] [Medline: 30346819]
33. Varney JE, Weiland TJ, Inder WJ, Jelinek GA. Effect of hospital-based telephone coaching on glycaemic control and adherence to management guidelines in type 2 diabetes, a randomised controlled trial. Intern Med J 2014 Sep;44(9):890-897. [doi: 10.1111/imj.12515] [Medline: 24963611]
34. Schechter CB, Walker EA, Ortega FM, Chamany S, Silver LD. Costs and effects of a telephonic diabetes self-management support intervention using health educators. J Diabetes Complications 2016 Mar;30(2):300-305 [FREE Full text] [doi: 10.1016/j.jdiacomp.2015.11.017] [Medline: 26750743]
35. Varney JE, Liew D, Weiland TJ, Inder WJ, Jelinek GA. The cost-effectiveness of hospital-based telephone coaching for people with type 2 diabetes: a 10 year modelling analysis. BMC Health Serv Res 2016 Sep 27;16(1):521 [FREE Full text] [doi: 10.1186/s12913-016-1645-6] [Medline: 27678079]
36. Cho JH, Kim H, Yoo SH, Jung CH, Lee WJ, Park CY, et al. An Internet-based health gateway device for interactive communication and automatic data uploading: Clinical efficacy for type 2 diabetes in a multi-centre trial. J Telemed Telecare 2017 Jul;23(6):595-604. [doi: 10.1177/1357633X16657500] [Medline: 27381040]
37. Egede LE, Williams JS, Voronca DC, Knapp RG, Fernandes JK. Randomized Controlled Trial of Technology-Assisted Case Management in Low Income Adults with Type 2 Diabetes. Diabetes Technol Ther 2017 Aug;19(8):476-482. [doi: 10.1089/dia.2017.0006] [Medline: 28581821]
38. Tildesley HD, Mazanderani AB, Ross SA. Effect of Internet therapeutic intervention on A1C levels in patients with type 2 diabetes treated with insulin. Diabetes Care 2010 Aug;33(8):1738-1740 [FREE Full text] [doi: 10.2337/dc09-2256] [Medline: 20668152]

39. Wild SH, Hanley J, Lewis SC, McKnight JA, McCloughan LB, Padfield PL, et al. Supported Telemonitoring and Glycemic Control in People with Type 2 Diabetes: The Telescot Diabetes Pragmatic Multicenter Randomized Controlled Trial. PLoS Med 2016 Jul;13(7):e1002098 [FREE Full text] [doi: 10.1371/journal.pmed.1002098] [Medline: 27458809]
40. Wakefield BJ, Holman JE, Ray A, Scherubel M, Adams MR, Hillis SL, et al. Effectiveness of home telehealth in comorbid diabetes and hypertension: a randomized, controlled trial. Telemed J E Health 2011 May;17(4):254-261. [doi: 10.1089/tmj.2010.0176] [Medline: 21476945]
41. Luley C, Blaik A, Reschke K, Klose S, Westphal S. Weight loss in obese patients with type 2 diabetes: effects of telemonitoring plus a diet combination - the Active Body Control (ABC) Program. Diabetes Res Clin Pract 2011 Mar;91(3):286-292. [doi: 10.1016/j.diabres.2010.11.020] [Medline: 21168231]
42. Goodarzi M, Ebrahimzadeh I, Rabi A, Saedipoor B, Jafarabadi MA. Impact of distance education via mobile phone text messaging on knowledge, attitude, practice and self efficacy of patients with type 2 diabetes mellitus in Iran. J Diabetes Metab Disord 2012;11(1):10 [FREE Full text] [doi: 10.1186/2251-6581-11-10] [Medline: 23497632]
43. Lim S, Kang SM, Kim KM, Moon JH, Choi SH, Hwang H, et al. Multifactorial intervention in diabetes care using real-time monitoring and tailored feedback in type 2 diabetes. Acta Diabetol 2016;53(2):189-198. [doi: 10.1007/s00592-015-0754-8] [Medline: 25936739]
44. Fang R, Deng X. Electronic messaging intervention for management of cardiovascular risk factors in type 2 diabetes mellitus: A randomised controlled trial. J Clin Nurs 2018 Feb;27(3-4):612-620. [doi: 10.1111/jocn.13962] [Medline: 28700102]
45. Fortmann AL, Gallo LC, Garcia MI, Taleb M, Euyoque JA, Clark T, et al. Dulce Digital: An mHealth SMS-Based Intervention Improves Glycemic Control in Hispanics With Type 2 Diabetes. Diabetes Care 2017 Oct;40(10):1349-1355. [doi: 10.2337/dc17-0230] [Medline: 28600309]
46. Kim H, Sun C, Yang SJ, Sun L, Li F, Choi IY, et al. Randomized, Open-Label, Parallel Group Study to Evaluate the Effect of Internet-Based Glucose Management System on Subjects with Diabetes in China. Telemed J E Health 2016 Aug;22(8):666-674. [doi: 10.1089/tmj.2015.0170] [Medline: 26938489]
47. Avdal EU, Kizilci S, Demirel N. The effects of web-based diabetes education on diabetes care results: a randomized control study. Comput Inform Nurs 2011 Feb;29(2):101-106. [doi: 10.1097/NCN.0b013e3181fcbdc6] [Medline: 21099675]
48. Cho J, Kwon H, Kim H, Oh J, Yoon K. Effects on diabetes management of a health-care provider mediated, remote coaching system via a PDA-type glucometer and the Internet. J Telemed Telecare 2011;17(7):365-370. [doi: 10.1258/jtt.2011.100913] [Medline: 21933896]
49. Kim S, Kim H. Effectiveness of mobile and internet intervention in patients with obese type 2 diabetes. Int J Med Inform 2008 Jun;77(6):399-404. [doi: 10.1016/j.ijmedinf.2007.07.006] [Medline: 17881285]
50. Lee JY, Chan CKY, Chua SS, Ng CJ, Paraidathathu T, Lee KKC, et al. Telemonitoring and Team-Based Management of Glycemic Control on People with Type 2 Diabetes: a Cluster-Randomized Controlled Trial. J Gen Intern Med 2020 Jan;35(1):87-94. [doi: 10.1007/s11606-019-05316-9] [Medline: 31512187]
51. Dario C, Toffanin R, Calcaterra F, Saccavini C, Stafylas P, Mancin S, et al. Telemonitoring of Type 2 Diabetes Mellitus in Italy. Telemed J E Health 2017;23(2):143-152. [doi: 10.1089/tmj.2015.0224] [Medline: 27379995]
52. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014 Jun;63(6):745-54.e6. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
53. Burner E, Lam CN, DeRoss R, Kagawa-Singer M, Menchine M, Arora S. Using Mobile Health to Improve Social Support for Low-Income Latino Patients with Diabetes: A Mixed-Methods Analysis of the Feasibility Trial of TExT-MED + FANS. Diabetes Technol Ther 2018 Dec;20(1):39-48. [doi: 10.1089/dia.2017.0198] [Medline: 29227155]
54. Del Prato S, Nicolucci A, Lovagnini-Scher AC, Turco S, Leotta S, Vespasiani G, ELEONOR Study Group. Telecare Provides comparable efficacy to conventional self-monitored blood glucose in patients with type 2 diabetes titrating one injection of insulin glulisine-the ELEONOR study. Diabetes Technol Ther 2012 Feb;14(2):175-182. [doi: 10.1089/dia.2011.0163] [Medline: 22013886]
55. Ramadas A, Chan CKY, Oldenburg B, Hussein Z, Quek KF. Randomised-controlled trial of a web-based dietary intervention for patients with type 2 diabetes: changes in health cognitions and glycemic control. BMC Public Health 2018 Dec 08;18(1):716 [FREE Full text] [doi: 10.1186/s12889-018-5640-1] [Medline: 29884161]
56. Cho J, Lee H, Lim D, Kwon H, Yoon K. Mobile communication using a mobile phone with a glucometer for glucose control in Type 2 patients with diabetes: as effective as an Internet-based glucose monitoring system. J Telemed Telecare 2009;15(2):77-82. [doi: 10.1258/jtt.2008.080412] [Medline: 19246607]
57. Cho J, Choi Y, Kim H, Lee J, Yoon K. Effectiveness and safety of a glucose data-filtering system with automatic response software to reduce the physician workload in managing type 2 diabetes. J Telemed Telecare 2011;17(5):257-262. [doi: 10.1258/jtt.2011.101006] [Medline: 21628421]
58. Jeong JY, Jeon J, Bae K, Choi Y, Park K, Kim J, et al. Smart Care Based on Telemonitoring and Telemedicine for Type 2 Diabetes Care: Multi-Center Randomized Controlled Trial. Telemed J E Health 2018 Aug;24(8):604-613. [doi: 10.1089/tmj.2017.0203] [Medline: 29341843]

59. Rodríguez-Idígoras MI, Sepúlveda-Muñoz J, Sánchez-Garrido-Escudero R, Martínez-González JL, Escolar-Castelló JL, Paniagua-Gómez IM, et al. Telemedicine influence on the follow-up of type 2 diabetes patients. Diabetes Technol Ther 2009 Jul;11(7):431-437. [doi: 10.1089/dia.2008.0114] [Medline: 19580356]
60. Pressman AR, Kinoshita L, Kirk S, Barbosa GM, Chou C, Minkoff J. A novel telemonitoring device for improving diabetes control: protocol and results from a randomized clinical trial. Telemed J E Health 2014 Feb;20(2):109-114. [doi: 10.1089/tmj.2013.0157] [Medline: 24404816]
61. Dienstl M, Kempf K, Schulz C, Kruse J, Martin S. Effect of Telemedicine on Glucometabolic Control and Quality of Life in Patients with Type 2 Diabetes Mellitus (in German). Diabetologie und Stoffwechsel 2011 Jun 21;6(03):164-169. [doi: 10.1055/s-0031-1271460]
62. Jha S, Dogra S, Yadav A, Siddiqui S, Panda M, Srivastava K, et al. A prospective observational study to assess the effectiveness of an electronic health (E-health) and mobile health (M-health) platform versus conventional care for the management of diabetes mellitus. Int J Diabetes Dev Ctries 2016 Jun 9;36(4):529-534. [doi: 10.1007/s13410-016-0501-x]
63. Kesavadev J, Shankar A, Pillai PBS, Krishnan G, Jothydev S. Cost-effective use of telemedicine and self-monitoring of blood glucose via Diabetes Tele Management System (DTMS) to achieve target glycosylated hemoglobin values without serious symptomatic hypoglycemia in 1,000 subjects with type 2 diabetes mellitus--a retrospective study. Diabetes Technol Ther 2012 Sep;14(9):772-776. [doi: 10.1089/dia.2012.0088] [Medline: 22734662]
64. Zhou P, Xu L, Liu X, Huang J, Xu W, Chen W. Web-based telemedicine for management of type 2 diabetes through glucose uploads: a randomized controlled trial. Int J Clin Exp Pathol 2014;7(12):8848-8854 [FREE Full text] [Medline: 25674254]
65. Crowley MJ, Edelman D, McAndrew AT, Kistler S, Danus S, Webb JA, et al. Practical Telemedicine for Veterans with Persistently Poor Diabetes Control: A Randomized Pilot Trial. Telemed J E Health 2016 May;22(5):376-384. [doi: 10.1089/tmj.2015.0145] [Medline: 26540163]
66. Liou J, Soon M, Chen C, Huang T, Chen Y, Yeh Y, et al. Shared care combined with telecare improves glycemic control of diabetic patients in a rural underserved community. Telemed J E Health 2014 Feb;20(2):175-178. [doi: 10.1089/tmj.2013.0037] [Medline: 24320193]
67. Wakefield BJ, Koopman RJ, Keplinger LE, Bomar M, Bernt B, Johanning JL, et al. Effect of home telemonitoring on glycemic and blood pressure control in primary care clinic patients with diabetes. Telemed J E Health 2014 Mar;20(3):199-205 [FREE Full text] [doi: 10.1089/tmj.2013.0151] [Medline: 24404819]
68. Stone RA, Rao RH, Sevick MA, Cheng C, Hough LJ, Macpherson DS, et al. Active care management supported by home telemonitoring in veterans with type 2 diabetes: the DiaTel randomized controlled trial. Diabetes Care 2010 Mar;33(3):478-484 [FREE Full text] [doi: 10.2337/dc09-1012] [Medline: 20009091]
69. Nicolucci A, Cercone S, Chiriatti A, Muscas F, Gensini G. A Randomized Trial on Home Telemonitoring for the Management of Metabolic and Cardiovascular Risk in Patients with Type 2 Diabetes. Diabetes Technol Ther 2015 Aug;17(8):563-570. [doi: 10.1089/dia.2014.0355] [Medline: 26154338]
70. Kempf K, Altpeter B, Berger J, Reuß O, Fuchs M, Schneider M, et al. Efficacy of the Telemedical Lifestyle intervention Program TeLiPro in Advanced Stages of Type 2 Diabetes: A Randomized Controlled Trial. Diabetes Care 2017 Dec;40(7):863-871. [doi: 10.2337/dc17-0303] [Medline: 28500214]
71. Stone RA, Sevick MA, Rao RH, Macpherson DS, Cheng C, Kim S, et al. The Diabetes Telemonitoring Study Extension: an exploratory randomized comparison of alternative interventions to maintain glycemic control after withdrawal of diabetes home telemonitoring. J Am Med Inform Assoc 2012;19(6):973-979 [FREE Full text] [doi: 10.1136/amiajnl-2012-000815] [Medline: 22610495]
72. Hsu WC, Lau KHK, Huang R, Ghiloni S, Le H, Gilroy S, et al. Utilization of a Cloud-Based Diabetes Management Program for Insulin Initiation and Titration Enables Collaborative Decision Making Between Healthcare Providers and Patients. Diabetes Technol Ther 2016 Feb;18(2):59-67 [FREE Full text] [doi: 10.1089/dia.2015.0160] [Medline: 26645932]
73. von Storch K, Graaf E, Wunderlich M, Rietz C, Polidori MC, Woopen C. Telemedicine-Assisted Self-Management Program for Type 2 Diabetes Patients. Diabetes Technol Ther 2019 Sep;21(9):514-521. [doi: 10.1089/dia.2019.0056] [Medline: 31287736]
74. Carter EL, Nunlee-Bland G, Callender C. A patient-centric, provider-assisted diabetes telehealth self-management intervention for urban minorities. Perspect Health Inf Manag 2011;8(Winter):1b [FREE Full text] [Medline: 21307985]
75. Warren R, Carlisle K, Mihala G, Scuffham PA. Effects of telemonitoring on glycaemic control and healthcare costs in type 2 diabetes: A randomised controlled trial. J Telemed Telecare 2018;24(9):586-595. [doi: 10.1177/1357633X17723943] [Medline: 28814128]
76. Tang PC, Overhage JM, Chan AS, Brown NL, Aghighi B, Entwistle MP, et al. Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc 2013 May 1;20(3):526-534 [FREE Full text] [doi: 10.1136/amiajnl-2012-001263] [Medline: 23171659]
77. Greenwood DA, Blozis SA, Young HM, Nesbitt TS, Quinn CC. Overcoming Clinical Inertia: A Randomized Clinical Trial of a Telehealth Remote Monitoring Intervention Using Paired Glucose Testing in Adults With Type 2 Diabetes. J Med Internet Res 2015;17(7):e178 [FREE Full text] [doi: 10.2196/jmir.4112] [Medline: 26199142]

78. Steventon A, Bardsley M, Doll H, Tuckey E, Newman SP. Effect of telehealth on glycaemic control: analysis of patients with type 2 diabetes in the Whole Systems Demonstrator cluster randomised trial. BMC Health Serv Res 2014;14:334 [FREE Full text] [doi: 10.1186/1472-6963-14-334] [Medline: 25100190]
79. Su D, Zhou J, Kelley MS, Michaud TL, Siahpush M, Kim J, et al. Does telemedicine improve treatment outcomes for diabetes? A meta-analysis of results from 55 randomized controlled trials. Diabetes Res Clin Pract 2016 Jun;116:136-148. [doi: 10.1016/j.diabres.2016.04.019] [Medline: 27321329]

Abbreviations

BP: blood pressure
CT: clinical trial
DRQoL: diabetes-related quality of life
FBG: fasting blood glucose
HbA1c: glycated hemoglobin A1c
HRQoL: health-related quality of life
IDF: International Diabetes Federation
MA: meta-analysis
MD: mean difference
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SR: systematic review
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 05.08.20; peer-reviewed by D Kim, K Lee, KC Wong; comments to author 21.09.20; revised version received 13.10.20; accepted 12.12.20; published 17.02.21.

Please cite as:
Eberle C, Stichling S
Effect of Telemetric Interventions on Glycated Hemoglobin A1c and Management of Type 2 Diabetes Mellitus: Systematic Meta-Review
J Med Internet Res 2021;23(2):e23252
URL: http://www.jmir.org/2021/2/e23252/
doi:10.2196/23252
PMID:33595447

Review

Clinical Improvements by Telemedicine Interventions Managing Type 1 and Type 2 Diabetes: Systematic Meta-review

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

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

Corresponding Author:
Claudia Eberle, MD, Prof Dr
Medicine with specialization in Internal Medicine and General Medicine
Hochschule Fulda - University of Applied Sciences
Leipziger Strasse 123
Fulda, 36037
Germany
Phone: 49 661 9640 ext 6328
Fax: 49 661 9640 649
Email: claudia.eberle@hs-fulda.de

Abstract

Background: Diabetes mellitus (DM) is one of the world's greatest health threats with rising prevalence. Global digitalization leads to new digital approaches in diabetes management, such as telemedical interventions. Telemedicine, which is the use of information and communication technologies, may provide medical services over spatial distances to improve clinical patient outcomes by increasing access to diabetes care and medical information.

Objective: This study aims to examine whether telemedical interventions effectively improve diabetes control using studies that pooled patients with type 1 diabetes mellitus (T1DM) and type 2 diabetes mellitus (T2DM), and whether the benefits are greater in patients diagnosed with T2DM than in those diagnosed with T1DM. We analyzed the primary outcome glycated hemoglobin A_{1c} (HbA_{1c}) and the secondary outcomes fasting blood glucose (FBG), blood pressure (BP), body weight, BMI, quality of life (QoL), cost, and time saving.

Methods: Publications were systematically identified by searching Cochrane Library, MEDLINE via PubMed, Web of Science Core Collection, Embase, and CINAHL databases for studies published between January 2008 and April 2020, considering systematic reviews (SRs), meta-analyses (MAs), randomized controlled trials (RCTs), and clinical trials (CTs). Study quality was assessed using the *A Measurement Tool to Assess Systematic Reviews*, *Effective Public Health Practice Project*, and *National Institute for Health and Care Excellence* qualitative checklist. We organized the trials by communication technologies in *real-time video or audio interventions*, *asynchronous interventions*, and *combined interventions* (synchronous and asynchronous communication).

Results: From 1116 unique citations, we identified 31 eligible studies (n=15 high, n=14 moderate, n=1 weak, and n=1 critically low quality). We selected 21 SRs and MAs, 8 RCTs, 1 non-RCT, and 1 qualitative study. Of the 10 trials, 3 were categorized as *real-time video*, 1 as *real-time video and audio*, 4 as *asynchronous*, and 2 as *combined* intervention. Significant decline in HbA_{1c} levels based on pooled T1DM and T2DM patients data ranged from −0.22% weighted mean difference (WMD; 95% CI −0.28 to −0.15; $P<.001$) to −0.64% mean difference (95% CI −1.01 to −0.26; $P<.001$). The intervention effect on lowering HbA_{1c} values might be significantly smaller for patients with T1DM than for patients with T2DM. Evidence on the impact on BP, body weight, FBG, cost effectiveness, and time saving was smaller compared with HbA_{1c} but indicated potential in some publications.

Conclusions: Telemedical interventions might be clinically effective in improving diabetes control overall, and they might significantly improve HbA_{1c} concentrations. Patients with T2DM could benefit more than patients with T1DM regarding lowering HbA_{1c} levels. Further studies with longer duration and larger cohorts are necessary.

(J Med Internet Res 2021;23(2):e23244) doi:10.2196/23244

KEYWORDS

type 1 diabetes; type 2 diabetes; eHealth; telemedicine; disease management; systematic review; mobile phone

Introduction

Background

Diabetes mellitus (DM) is one of the world's greatest health threats [1]. The global prevalence of DM will increase from around 463 million (2019) to approximately 700 million in 2045, that is, by 51% [2]. DM is a chronic metabolic disorder associated with insulin resistance and hyperglycemia [1]. An increased blood sugar level can lead to long-term damage to the heart, eyes, nerves, kidneys, and blood vessels [3].

Type 1 diabetes mellitus (T1DM) is based on autoimmune beta cell destruction, leading to absolute insulin deficiency, whereas type 2 diabetes mellitus (T2DM) is based on a progressive loss of beta cell insulin secretion based on the background of insulin resistance [1]. A healthy diet, regular physical activity, medication, normal body weight, and blood glucose control are important therapy components to mitigate or delay consequences [3]. Access to affordable therapy is critical to the survival of diabetes patients [3]. Effective therapies are important for diabetes control. Patients with diabetes need to understand the disease and be actively involved in diabetes management for optimal therapeutic effects [4].

Global digitalization offers innovative digital opportunities for intensive diabetes management. Diabetes technology includes hardware, software, and technical devices that help to control the disease with regard to the therapy components mentioned [4]. Technological possibilities are constantly evolving and rapidly growing. Telemedical interventions in the context of diabetes management show the potential to effectively improve diabetes control [5]. Telemedicine, a term shaped in the 1970s, characterizes the "use of [information and communication technologies] to improve patient outcomes by increasing access to care and medical information" [6]. Telemedicine offers the following features: providing clinical support, overcoming geographical and physical limits, improving health-related outcomes, and applying information and communication technologies [6].

Telemedicine is a part of eHealth, defined as the use of information and communication technologies for health, which was developed in the 1990s [6]. Nowadays, the term *digital health* is recommended as it creates a link between digitalization, health, lifestyle, and community [7]. Telemedicine provides certain benefits, such as an improved access to health care services, an enhanced quality of health and care management, and potential cost and time efficiencies [8]. The model for assessment of telemedicine (MAST) provides a structured framework consisting of different domains for appraising the effectiveness of telemedicine interventions [9]. We have considered the following MAST domains as part of this research: (1) health problems and description of the application, (2) clinical effectiveness, (3) patient perspectives, (4) economic aspects, and (5) organizational aspects.

However, although lifestyle changes are part of both T1DM and T2DM management, the focus of T1DM is on insulin therapy and that of T2DM is on nutrition and regular exercise [10]. Telemedical diabetes management may be more effective in T2DM patients by addressing modifiable factors such as nutrition and exercise [10]. Furthermore, T1DM patients are younger, and the management of T1DM in childhood and adolescence is both a medical and psychosocial challenge, as adolescence is a very vulnerable period [11].

Objectives

We recently examined the clinical effectiveness of telemedical interventions for diabetes therapy in T1DM and T2DM patients separately. Tendency differences in the clinical effectiveness of telemedical interventions between these two types of diabetes require further research in greater detail. After these separate analyses, that is, on the one hand, the analysis of T1DM patients and on the other hand that of T2DM patients exclusively, we want to focus on publications that include both types of diabetes. Therefore, in this systematic meta-review, we examined 2 hypotheses using studies that pooled T1DM and T2DM participants. We hypothesize that (1) telemedical interventions effectively improve diabetes control overall and (2) the benefits may be significantly greater in patients diagnosed with T2DM than in patients diagnosed with T1DM. As part of the first hypothesis, we also aimed to identify which intervention types (technologies) are particularly successful and effective. The focus of this systematic meta-review is on the communication between patients and health care professionals. As part of the second hypothesis, we focused on studies that specifically investigated differences between T1DM and T2DM patients with regard to the clinical effectiveness of telemedical interventions. In general, we concentrated on the following clinically relevant outcomes: primary outcome glycated hemoglobin A_{1c} (HbA_{1c}) and secondary outcomes fasting blood glucose (FBG), blood pressure (BP), body weight, BMI, quality of life (QoL), cost, and time saving. Furthermore, we included randomized controlled trials (RCTs), systematic reviews (SRs), meta-analyses (MAs), and clinical trials (CTs) for a comprehensive analysis. This systematic meta-review was based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [12].

Methods

Search Strategy

Publications were systematically identified by searching the Cochrane Library, MEDLINE via PubMed, Web of Science Core Collection, EMBASE, and CINAHL databases for studies published between January 2008 and April 2020. First, we carried out a comprehensive literature search targeting T1DM, T2DM, gestational diabetes mellitus (GDM), and mixed studies involving different types of diabetes.

Search was conducted using the following keywords: ("diabetes mellitus") AND ("telemedicine" OR "telemonitoring" OR "telemedicine"). Medical Subject Headings and Embase Subject Headings terms as well as title/abstract terms were used. This was supplemented with a manual search of the reference lists. Ultimately, studies that included both T1DM and T2DM patients were selected for inclusion in this systematic meta-review. The search strategies are shown in Multimedia Appendix 1. Studies

were screened and selected by 2 independent reviewers, and any disagreement was resolved through consensus.

Eligibility Criteria

Studies that met the following inclusion criteria were selected: peer-reviewed studies with full text; including T1DM and T2DM patients; addressing telemedical interventions for diabetes management; published in English or German; participants transfer data and receive appropriate feedback from health care professionals; and study designs: RCTs, SRs, MAs, and CTs. Qualitative and quantitative studies were considered. Telemedicine was defined as "remote acquisition, recording and transmission of patient data via a telecommunications system to a health care provider for analysis and decision making" [8]. Therefore, videoconferences, telephone calls, asynchronous communication by emails, SMS text messaging, internet/web-based platforms, and mixed forms (eg, videoconferences and emails) were included.

On the basis of this, the following studies were excluded: poster, comments, letters, study protocols, proceedings papers, studies that did not specify the types of diabetes of the population, studies providing pooled data with patients diagnosed with GDM, studies providing description of technologies only; studies conducted on smartphone/mobile apps (we analyzed these separately in another research paper because of the different nature of the technology), pooled data with other technologies, duplicates, and papers focusing on prevention or diagnosis.

Data Extraction

Study characteristics (year of publication, study region, and design), patient characteristics (type of diabetes), intervention details (outcome measures, duration, intervention and control group, sample size, and information and communication technologies used), and main results and author's conclusions were extracted.

Outcomes of Interest

The primary outcome was HbA_{1c}, and the secondary outcomes were FBG, BP, body weight, BMI, health-related quality of life (HRQoL), diabetes-related quality of life (DRQoL), cost effectiveness, and time saving.

Data Synthesis and Analysis

A previously developed scheme to structure the constituent studies according to intervention types was used for data synthesis and analysis. On the basis of the information and communication technologies used between health care professionals and diabetes patients, we chose 4 types of intervention: *real-time video interventions*, *real-time audio interventions*, *asynchronous interventions*, and *combined interventions*.

Real-time video encompasses synchronous communication that takes place face to face, whereas *real-time audio* also covers synchronous communication, which occurs via telephone calls. *Asynchronous* depicts time-shifted communication, for example, using emails, server, SMS text messaging, and internet-/web-based platforms. The last category *combined interventions* combines *real-time* and *asynchronous* communication and accordingly includes elements from both forms.

Assessment of Risk of Bias

The quality appraisal of the studies for the assessment of the risks of bias was carried out using 3 different validated instruments, as we included different study designs. For SRs and MAs, we applied *A MeaSurement Tool to Assess systematic Reviews* (AMSTAR 2). For RCTs and CTs, the *Effective Public Health Practice Project* (EPHPP) was applied. Moreover, the *National Institute for Health and Care Excellence* (NICE) quality appraisal checklist was deployed for qualitative studies. The tools grade the study quality into (critically) low/weak, moderate, and high/strong.

Results

Characteristics of the Included Studies

A systematic literature search identified 1647 citations. After duplicates were removed, we screened 1116 citations and excluded 875 ineligible papers based on their titles and abstracts. After assessing 241 studies with full text and excluding 72 inappropriate publications based on our inclusion and exclusion criteria, we manually searched reference lists and yielded 184 studies that covered T1DM, T2DM, and GDM. Finally, we included 31 suitable studies that comprised T1DM and T2DM patients in this systematic meta-review.

We identified 21 SRs and MAs, 8 RCTs, 1 non-RCT, and 1 qualitative study. SR and MA aside, 3 trials were conducted in the United States, 1 in Taiwan, 2 in the United Kingdom, 1 in Greece, 1 in Australia, and 1 in Israel.

In general, 8 SRs/MAs were rated with high, 12 with moderate and 1 with critically low quality based on AMSTAR 2. Furthermore, of 9 trials that were assessed using EPHPP, 6 were high/strong, 2 were moderate, and 1 was weak in quality. The qualitative study was assessed using the NICE checklist for qualitative studies, which resulted in high (++) quality.

Of the 10 identified intervention studies (without SRs and MAs), 3 were categorized as *real-time video interventions*, 1 as *real-time video and audio intervention*, 4 as *asynchronous interventions*, and 2 as *combined interventions* (real-time and asynchronous communication). Interestingly, none of the studies examined pure *real-time audio intervention*.

Overview of Presentation of Findings

The characteristics of the publications are summarized in Table 1 and Table 2. The search and selection protocol is allocated as a PRISMA flowchart in Multimedia Appendix 2 [12]. A summary of the studies, their characteristics, intervention details, main findings, and conclusions are provided in Multimedia Appendix 3 [5,13-42]. All studies, except for the qualitative study, were in a controlled design. Multimedia Appendix 4 summarizes the quality assessments using different tools.

Table 1. Baseline characteristics of all studies (n=31).

Characteristics	Value, n (%)
Study design	
SRs[a] & MAs[b]	21 (67)
RCT[c]	8 (25)
Non-RCT	1 (3)
Qualitative study	1 (3)
Year	
2008-2011	3 (9)
2012-2014	12 (38)
2015-2017	9 (29)
2018-2020	6 (19)

[a]SR: systematic review.

[b]MA: meta-analysis.

[c]RCT: randomized controlled trial.

Table 2. Baseline characteristics of the individual studies (n=10, without systematic reviews and meta-analysis).

Characteristics	Individual studies, n (%)
Location	
United States	3 (30)
Taiwan	1 (10)
United Kingdom	2 (20)
Greece	1 (10)
Australia	1 (10)
Israel	1 (10)
Intervention	
Real-time video	3 (30)
Real-time video+audio	1 (10)
Asynchronous	4 (40)
Combined (real-time+asynchronous)	2 (20)

Effects on Primary and Secondary Clinical Outcomes

We synthesized the effects on the clinical outcomes in Multimedia Appendix 5. Table 3 provides a summary of the *significant* intervention effects on the clinical outcomes (intra- and intergroup effects are listed). In some cases, positive effects were not significant, and some studies indicated obvious improvements, but *P* values were not available. The denominators do not add up to the total “n” (left-hand panel) because not every study examined only one outcome.

Table 3. Significant effects on primary outcome HbA_{1c} and secondary outcomes (intra- and intergroup).

Study/outcome	HbA_{1c}[a]	FBG[b]	BP[c]	Body weight	BMI	DRQoL[d]	HRQoL[e]	Cost	Time-saving	N/S[f] or significance not available
SR[g] & MA[h] (n=21)	(13/17)[i] +[j]	—[k]	(2/4) +	—	—	—	—	(2/5) +	N/A	—
Real-time video (n=3)	—	—	—	—	—	—	—	—	—	✓[l]
Real-time audio+video (n=1)	—	—	—	—	—	—	—	—	—	✓
Asynchronous (n=4)	(1/3) +	—	(1/1) +	—	(1/1) +	—	—	—	—	—
Combined (n=2)	—	—	—	(1/1) +	N/A	(1/1) +	—	—	—	—

[a]HbA_{1c}: glycated hemoglobin A_{1c}.

[b]FBG: fasting blood glucose.

[c]BP: blood pressure.

[d]DRQoL: diabetes-related quality of life.

[e]HRQoL: health-related quality of life.

[f]N/S: not significant.

[g]SR: systematic review.

[h]MA: meta-analysis.

[i]Number of publications with significant intervention effects versus number of all studies (inclusive significant effects that were not long term).

[j]+: outcome improvement.

[k]Missing data.

[l]N/S or significance not available.

SR and MA (n=21)

HbA_{1c} (n=17)

In general, all investigated SRs and MAs reported positive effects on HbA_{1c} values [5,13-28]. Most studies (n=13, 76.5%; high quality and moderate quality) showed significant improvements in HbA_{1c} concentrations. Significant reductions based on pooled T1DM and T2DM patients data ranged from −0.22% weighted mean difference (WMD; 95% CI −0.28 to −0.15; *P*<.001) by Wu et al [21] (moderate quality) to −0.64% mean difference (95% CI −1.01 to −0.26; *P*<.001) by So and Chung [17] (moderate quality). According to Faruque et al [13] (high quality), the intervention effect on HbA_{1c} was highest in studies with high baseline HbA_{1c} values and in interventions with asynchronous communication via web portals or text messaging. Toma et al [20] (high quality) also indicated the greatest significant reduction in asynchronous *internet only interventions* (−0.51%; 95% CI −0.68 to −0.34; *P*<.001). Su et al [18] (high quality) concluded that interventions with a duration of 6 months or less showed a greater decline in HbA_{1c} values (Hedge g=−0.56%; *P*<.001). Tao and Or [24] (high quality) also reported the greatest improvement in short-term interventions lasting 3 months or less (−0.54%, 95% CI −0.80 to −0.28; *P*<.001).

BP (n=4)

Overall, all studies (moderate quality and high quality) mentioned positive effects on BP [15,20,21,26]. Wu et al [21] (moderate quality) and Toma et al [20] (high quality) reported significant decreases in systolic and diastolic BP values compared with usual care: systolic BP WMD −1.92 mm Hg (95% CI −2.49 to −1.34; *P*<.001), diastolic BP WMD −1.31 mm Hg (95% CI −2.39 to −0.23; *P*<.001) [21], systolic BP −3.47 mm Hg (95% CI −5.0 to −1.94; *P*<.001), and diastolic BP −1.84 mm Hg (95% CI −2.98 to −0.70; *P*=.11) [20].

Body Weight (n=1)

Jong et al [26] (moderate quality) described that one telemedical intervention had positive effects on patients' body weight, but no significance was reported.

BMI (n=3)

Studies by Wu et al [21] (moderate quality), Hu et al [14] (moderate quality), and Marcolino et al [15] (moderate quality) found positive effects of telemedical interventions on BMI, but were not statistically significant. Wu et al [21] (moderate quality) reported a difference between the telehealth and usual care groups in controlling BMI (WMD=−0.14 kg/m² (95% CI −1.13 to 0.68; *P*=.79). Hu et al [14] (moderate quality) outlined a reduction of 0.27 kg/m² (95% CI 0.86 to −0.31; I^2=40%; *P*=.35).

DRQoL and HRQoL (n=3)

Two studies (66.7%) by Polisena et al [16] (high quality) and Wu et al [21] (moderate quality) concluded overall positive effects on DRQoL as well as HRQoL (not significant [21] and significance was not reported [16]), whereas Faruque et al [13] (high quality) found no effects on QoL.

Cost Effectiveness (n=5)

According to 4 papers (1 high quality, 2 moderate quality, and 1 critically low quality; 80.0%), telemedical interventions can be viewed cost effectively [5,23,29,30]. Tchero et al [5] (high quality) noted an incremental cost-effectiveness ratio (ICER)

in 3 studies of US $490, US $29,869, and US $464, per capita for each unit reduction in HbA_{1c}. Lee and Lee (moderate quality) [29] found a moderate cost-effectiveness in 7 telephone interventions of US $4744.32- US $86,276.50/quality-adjusted life year (ICER). Only Teljeur et al [31] (moderate quality) reported that telemedicine was not cost effective at all (based on 3 studies).

Barriers and Enablers (n=1)

MacDonald et al [32] (moderate quality) identified poorly designed interfaces as barriers and highly automated data entry and transmission, support by health care professionals and family, integration of users in the design process, and reliable technology as enablers for the implementation of communication technologies in diabetes management.

Real-Time Video Interventions (n=3)

HbA_{1c} (n=2)

The trials by Sood et al [33] (moderate quality) and Kearns et al [34] (weak quality) found no significant improvements in HbA_{1c} values (P>.05). Sood et al [33] (weekly video conferences) reported a decrease in HbA_{1c} (intervention −1.01% vs usual care −0.68%; P=.19).

BP (n=1)

Sood et al [33] (moderate quality) reported slightly, not significantly, increased BP values (+3.8 mm Hg) in the intervention group (P=.02).

Process Analysis of Video Consultations (n=1)

Fatehi et al [35] (high quality) analyzed video consultations qualitatively and found that health care professionals were confident with their feedback for diabetes patients via videoconferencing.

Real-time Audio+Video Intervention (n=1)

HRQoL (n=1)

According to a high quality study by Young et al [36] (combined telephone and videoconferencing intervention), physical and mental health clearly improved in the intervention group compared with usual care, but the difference was not statistically significant (P<.05).

Asynchronous Interventions (n=4)

HbA_{1c} (n=3)

In general, all RCTs (moderate quality and high quality) found reductions in HbA_{1c} values [37-39]. In addition, 2 RCTs reported significant improvements: Chen et al [37] (moderate quality) noted a significant improvement in the intervention group (P=.02), and Fountoulakis et al [38] (high quality) outlined a significant reduction in the intervention group at 3 (7.1±1.0%; P<.001) and 6 months (6.9±0.9%; P<.001), compared with the control group.

BP (n=1)

In a high quality RCT by Earle et al [40], systolic BP fell significantly in the intervention group (−6.5 mm Hg; 95% CI −0.8 to −12.2; P=.03) and remained unchanged in the control group (P=.57).

BMI (n=1)

According to Fountoulakis et al [38] (high quality), a significant BMI reduction (P<.05) was observed in the intervention participants at 6 months as well as at 6-months off telemonitoring compared with baseline, whereas intergroup analysis was not significant.

Combined Interventions (n=2)

HbA_{1c} (n=2)

In 2 high quality RCTs by Leichter et al [42] and Boaz et al [41], the authors observed no significant differences in post treatment HbA_{1c}, and the values slightly increased in the intervention groups.

FBG (n=1)

Boaz et al [41] (high quality) observed a nonsignificant decrease in FBG values compared with the control group.

BP (n=1)

Leichter et al [42] (high quality) reported a nonsignificant increase in systolic BP in the intervention group.

Body Weight (n=2)

Although Leichter et al [42] (high quality) showed a significantly greater reduction in body weight in the intervention group (−5.2 vs −0.7 pounds; P=.04), Boaz et al [41] (high quality) found a slight, not significant, weight gain in treatment subjects.

BMI (n=1)

According to Leichter et al [42] (high quality), BMI decreased in the intervention group, with no significant between-group differences at 12 months [42].

DRQoL (n=1)

Boaz et al [41] (high quality) reported that intervention patients reported significantly greater posttreatment experiences of improved QoL [41]. The DRQoL measures *being clinically symptom-free* (71% vs 11%; P=.003), *having no hypoglycemic events* (82% vs 17%; P<.001), and *having no hyperglycemic events* (65% vs 17%; P=.004) were significantly more frequent in the intervention group.

Time Saving (n=1)

According to Leichter et al [42] (high quality), the clinician time requirements for intervention subjects were reduced by 40% (significance not reported).

Differences Between T1DM and T2DM

In total, 5 studies (3 high quality and 2 moderate quality) specifically examined differences in treatment effects between T1DM and T2DM patients [5,18,20,22,27]. Of these, 4 studies [5,18,20,27] reported smaller effects for T1DM, and 1 publication by Hanlon et al [22] (moderate quality) observed significantly improved glycemic control (HbA_{1c}) in T2DM but not T1DM.

Su et al [18] (high quality) described that telemedicine was most effective in T2DM patients (Hedge g=−0.63; *P*<.001), whereas the effect was smaller for T1DM patients (Hedge g=−0.27; *P*=.03) and for T1DM and T2DM combined (Hedge g=−0.34; *P*=.003). The difference was statistically significant between T1DM and T2DM (Q statistics=4.25; *P*=.04). Tchero et al [5] (high quality) observed similar findings. Patients diagnosed with T2DM experienced a significantly greater reduction in HbA_{1c} concentrations compared with those diagnosed with T1DM (Hedge g=−0.48, *P*=.001 vs −0.26, *P*=.05; Q=1935.75, *P*<.001). In accordance with this, Toma et al [20] (high quality) reported a significantly smaller effect in T1DM (−0.12%; 95% CI −0.32 to −0.08; *P*=.26) than in T2DM (−0.55%; 95% CI −0.68 to −0.42; *P*<.001). Kitsiou et al [27] (moderate quality) outlined similar findings, but this finding was not significant. Telemedical interventions improved glycemic control (HbA_{1c}) compared with usual care: mean difference −0.8% (95% CI −1.11. to 0.5; n=280 patients; *P*<.05) for patients with T2DM and mean difference −0.3% (95% CI 0.0 to 0.5; n=645 patients; *P*>.05) for patients with T1DM.

Discussion

Principal Findings

In total, telemedical interventions might improve diabetes management in studies that pooled T1DM and T2DM participants overall and probably improve HbA_{1c} values, which corresponds to our first hypothesis. T2DM patients could benefit more from telemedical interventions than T1DM patients regarding lowering HbA_{1c} levels, but limitations must be taken into account, such as higher baseline HbA_{1c} values, short duration, small effect sizes, and partially small cohorts.

T1DM and T2DM have contrasting pathogenesis, which also results in appropriate therapy recommendations with different priorities [1]. Patients with T1DM are insulin deficient and require insulin therapy, and patients with T2DM are insulin resistant [1]. Although lifestyle changes are part of both T1DM and T2DM management, the focus of T1DM is on insulin therapy, whereas T2DM is, besides medication, more focused on nutrition regarding overweight and regular exercise [10]. Telemedical diabetes management may be more effective in T2DM patients by addressing modifiable factors such as nutrition and exercise [10]. Furthermore, T1DM patients are a unique group with special needs [11]. They are rather younger people, although both types of diabetes can occur at any age [1]. Management of T1DM in childhood and adolescence is both a medical and psychosocial challenge, as adolescence is a very vulnerable period [11]. For example, hormonal fluctuations and complications related to excessive insulin daily doses (eg, menstrual irregularities and weight gain) [43] as well as resistance to parents and health care services [11] play a role. These special features might have an impact on the differences between the types of diabetes. The findings further illustrate the importance of tailoring diabetes management to the patient and his/her needs [32].

Primary Outcome HbA_{1c}

HbA_{1c} was certainly the most investigated outcome. Numerous studies with high quality and moderate quality showed that telemedical interventions effectively reduced HbA_{1c} values significantly, up to −0.64% mean difference (95% CI −1.01 to −0.26; *P*<.001) by So and Chung [17]. Few *real-time video interventions* with moderate quality and WEAK QUALITY indicated clear but not significant improvements in HbA_{1c} levels. The *asynchronous interventions* showed significant HbA_{1c} reductions in some moderate quality and high quality studies. In contrast, few *combined interventions* generally miscarried to reduce HbA_{1c} values effectively, but we only analyzed 2 appropriate trials.

Our findings regarding more effective HbA_{1c} improvements compared with usual care are an important success for digital diabetes therapy because strict glycemic control is critical for both T1DM and T2DM patients [1,44]. Optimal glycemic control reduces and prevents micro- and macrovascular diabetic complications [44]. Microvascular complications such as nephropathy, retinopathy, and neuropathy as well as macrovascular events, such as cardiovascular disease and stroke, are closely associated with T1DM and T2DM, respectively [1,45]. The *legacy effect* showed that an HbA_{1c} value in the target area is predominant at an early stage because increased HbA_{1c} concentrations in the first year after a T2DM diagnosis increase the risk of micro- and macrovascular events in the long term. Even patients with HbA_{1c} levels between 7% and <8% had an increased risk compared with patients with HbA_{1c} levels below 6.5%.

BP: Telemedical interventions improved BP values in several moderate quality and high quality studies. There were positive significant effects on BP through an *asynchronous* high quality intervention. Clinical studies have shown an association between T2DM, vascular disease, and the present risk factor hypertension [45]. Therefore, the observed positive tendencies should be investigated further.

Body weight: Telemedical interventions demonstrated positive tendencies in terms of body weight improvement, which is particularly evident in a moderate quality review, but the number of identified studies was rare. These positive tendencies should be examined more closely.

FBG: One *combined intervention* (synchronous and asynchronous communication) examined the outcome FBG and indicated a nonsignificant decrease in FBG values.

QoL: Several studies with high quality and moderate quality indicated positive effects of telemedicine on DRQoL and HRQoL. The high quality real-time audio and video intervention significantly improved physical and mental health. Likewise, the high quality combined intervention significantly enhanced DRQoL.

Cost: In general, 80% (4/5) of the SRs and MAs demonstrated cost effectiveness of telemedical interventions. Telemedicine appears to be a cost-effective option in the context of diabetes

management with limitations, as cost effectiveness is related to country conditions and settings.

Time saving: Although only 1 study examined this outcome, this high quality study showed that the clinician time could be reduced.

Overall, important outcomes with little evidence available, such as FBG, body weight, cost, and time saving in particular, all showed positive tendencies and should be investigated further.

Type of intervention: To determine the types of telemedical interventions that are particularly effective and successful, we previously developed a scheme that divides the interventions into 4 categories. Overall, only a few studies could be identified for these categories. Interestingly, we could not identify any pure *real-time audio interventions* (communication via telephone calls only), just 1 study that included both real-time video and audio communication. *Real-time video interventions* showed obvious improvements in HbA_{1c} values.*Asynchronous interventions* significantly improved HbA_{1c} levels, BP, and BMI. In addition, *combined interventions* showed potential in improving BMI and DRQoL. No type of intervention was clearly superior. Some SRs and MAs indicated that asynchronous interventions (web portal, email, and internet) were more successful. The type of intervention or the form of communication should be adapted to the needs of diabetes patients [32]. A poorly designed interface can be a barrier for the patient and impair the success of the digital intervention [32]. Reliable technologies and support by health care professionals and families promote the success of telemedical approaches [32]. When patients believe that the technology is useful and offers feedback and flexibility, they are more likely to adopt this new option [32]. In total, we assume that telemedical approaches improve compliance, empower patients, and enable more intensive therapy as well as effective disease management on the part of health care professionals and patients.

From a clinical point of view, it is important to know the effect size that only results from modifying the communication level by using telemedicine. In clinical practice, these effects must be known and added to the therapeutic effects (eg, by insulin). This is also important to provide evidence-based recommendations. In addition, the relatively short duration of the studies must be taken into account. In principle, further studies with longer duration and larger cohorts are necessary to analyze the effect sizes in more detail.

The strengths of this systematic meta-review lie in the scheme used to categorize the telemedical interventions, the consideration of numerous study designs, the analysis of several clinically important outcomes, and the investigation of specific differences in the intervention effects between T1DM and T2DM patients.

Limitations

Some limitations may have affected our findings. There were large variations in the types of telemedical technology used in the SRs and MAs, as telemedicine is a broad term with different definitions. The number of constituent studies in our 4 categories for the technologies used and their sample sizes were small for a number of outcomes measured. The trials comprised diverse interventional approaches, durations, and frequencies of contact with health care clinicians. Furthermore, many outcome measurements led to insignificant results, which may indicate methodological weaknesses. Some of the SRs and MAs that examined the differences between T1DM and T2DM indicated possible confounding factors that must be taken into account. A high baseline HbA_{1c} can lead to a larger reduction in HbA_{1c} during any kind of intervention.

Comparison With Prior Work

The findings of this systematic meta-review are overall consistent with other SRs and MAs on telemedicine for diabetes management. An effective and significant reduction in HbA_{1c} values in patients diagnosed with T1DM or T2DM through telemedicine was also reported by numerous other authors in evidence synthesis [5,13,46]. Wu et al [21] also demonstrated the positive impact of telemedical diabetes management on BP, BMI, and QoL in their MA on telehealth for managing diabetes. Tchero et al [5] (clinical effectiveness of telemedicine based on 42 RCTs) and Su et al [18] (impact of telemedicine on HbA_{1c}, including 55 RCTs) reported obviously smaller effects for T1DM patients on HbA_{1c} values than for T2DM patients.

Conclusions

Taken together, this systematic meta-review demonstrated that telemedical interventions might be clinically effective in the management of populations consisting of T1DM and T2DM patients. The intervention effect on lowering HbA_{1c} values may be smaller for T1DM patients than for T2DM patients. Although evidence on the impact on BP, body weight, FBG, cost effectiveness, and time saving is smaller compared with HbA_{1c}, potential is indicated in some publications. Although none of the intervention type we categorized—*real-time video/audio communication*, *asynchronous communication*, and *combined communication* (*real-time and asynchronous*) was superior, each type generally showed improvements in clinical diabetes management.

Further research is needed regarding the differences between T1DM and T2DM with regard to the improvement of BP, body weight, and FBG. Overall, improvement in FBG values should be investigated in more detail, as we only identified one study measuring this outcome. Our findings indicate that telemedical applications are promising in the context of diabetes therapy, but further studies with longer duration and larger cohorts are necessary.

Acknowledgments

This research was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), project number EB 440/4-1. Therefore, the authors would like to thank the DFG for the strong support of this research work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[PDF File (Adobe PDF File), 514 KB - jmir_v23i2e23244_app1.pdf]

Multimedia Appendix 2
Search and selection protocol as PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.
[PDF File (Adobe PDF File), 624 KB - jmir_v23i2e23244_app2.pdf]

Multimedia Appendix 3
Overview of studies included in systematic meta-review.
[PDF File (Adobe PDF File), 788 KB - jmir_v23i2e23244_app3.pdf]

Multimedia Appendix 4
Quality assessments.
[PDF File (Adobe PDF File), 499 KB - jmir_v23i2e23244_app4.pdf]

Multimedia Appendix 5
Effects on primary outcome HbA_{1c} and secondary outcomes.
[PDF File (Adobe PDF File), 786 KB - jmir_v23i2e23244_app5.pdf]

References

1. American Diabetes Association. 2. Classification and Diagnosis of Diabetes:. Diabetes Care 2020 Jan;43(Suppl 1):S14-S31. [doi: 10.2337/dc20-S002] [Medline: 31862745]
2. International Diabetes Federation. IDF Diabetes Atlas 9th Edition. IDF Diabetes Atlas. 2020. URL: https://www.diabetesatlas.org/en/resources/ [accessed 2021-01-19]
3. World Health Organization. Diabetes. https://www.who.int. 2018. URL: https://www.who.int/news-room/fact-sheets/detail/diabetes [accessed 2021-01-19]
4. American Diabetes Association. 7. Diabetes technology: standards of medical care in diabetes-2020. Diabetes Care 2020 Jan;43(Suppl 1):S77-S88. [doi: 10.2337/dc20-S007] [Medline: 31862750]
5. Tchero H, Kangambega P, Briatte C, Brunet-Houdard S, Retali G, Rusch E. Clinical effectiveness of telemedicine in diabetes mellitus: a meta-analysis of 42 randomized controlled trials. Telemed J E Health 2019 Jul;25(7):569-583. [doi: 10.1089/tmj.2018.0128] [Medline: 30124394]
6. WHO Global Observatory for eHealth. Telemedicine: Opportunities and Developments in Member States: Report on the Second Global Survey on eHealth. Geneva: World Health Organization; 2010.
7. Meister S, Becker S, Leppert F, Drop L. Digital Health, Mobile Health und Co. – Wertschöpfung durch Digitalisierung und Datenverarbeitung. Wiesbaden: Springer Fachmedien Wiesbaden; 2016:185-212.
8. American Telemedicine Association (ATA). Telehealth: Defining 21st Century Care. Telehealth Basics - ATA. 2019. URL: https://www.americantelemed.org/resource/why-telemedicine/ [accessed 2021-01-19]
9. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: mast. Int J Technol Assess Health Care 2012 Jan;28(1):44-51. [doi: 10.1017/S0266462311000638] [Medline: 22617736]
10. Tan SY, Wong JL, Sim YJ, Wong SS, Elhassan SA, Tan SH, et al. Type 1 and 2 diabetes mellitus: a review on current treatment approach and gene therapy as potential intervention. Diabetes Metab Syndr 2019;13(1):364-372. [doi: 10.1016/j.dsx.2018.10.008] [Medline: 30641727]
11. Datz N, Müther S, Kordonouri O. Besonderheiten des Typ-1-diabetes im kindes- und jugendalter. Diabetologe 2019 May 24;15(4):353-364. [doi: 10.1007/s11428-019-0488-2]
12. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
13. Faruque LI, Wiebe N, Ehteshami-Afshar A, Liu Y, Dianati-Maleki N, Hemmelgarn BR, Alberta Kidney Disease Network. Effect of telemedicine on glycated hemoglobin in diabetes: a systematic review and meta-analysis of randomized trials. Can Med Assoc J 2017 Mar 6;189(9):E341-E364 [FREE Full text] [doi: 10.1503/cmaj.150885] [Medline: 27799615]

14. Hu Y, Wen X, Wang F, Yang D, Liu S, Li P, et al. Effect of telemedicine intervention on hypoglycaemia in diabetes patients: a systematic review and meta-analysis of randomised controlled trials. J Telemed Telecare 2019 Aug;25(7):402-413. [doi: 10.1177/1357633X18776823] [Medline: 29909748]
15. Marcolino MS, Maia JX, Alkmim MB, Boersma E, Ribeiro AL. Telemedicine application in the care of diabetes patients: systematic review and meta-analysis. PLoS One 2013;8(11):e79246 [FREE Full text] [doi: 10.1371/journal.pone.0079246] [Medline: 24250826]
16. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes Obes Metab 2009 Oct;11(10):913-930. [doi: 10.1111/j.1463-1326.2009.01057.x] [Medline: 19531058]
17. So CF, Chung JW. Telehealth for diabetes self-management in primary healthcare: a systematic review and meta-analysis. J Telemed Telecare 2018 Jun;24(5):356-364. [doi: 10.1177/1357633X17700552] [Medline: 28463033]
18. Su D, Zhou J, Kelley MS, Michaud TL, Siahpush M, Kim J, et al. Does telemedicine improve treatment outcomes for diabetes? A meta-analysis of results from 55 randomized controlled trials. Diabetes Res Clin Pract 2016 Jun;116:136-148. [doi: 10.1016/j.diabres.2016.04.019] [Medline: 27321329]
19. Su D, McBride C, Zhou J, Kelley MS. Does nutritional counseling in telemedicine improve treatment outcomes for diabetes? A systematic review and meta-analysis of results from 92 studies. J Telemed Telecare 2016 Sep;22(6):333-347. [doi: 10.1177/1357633X15608297] [Medline: 26442959]
20. Toma T, Athanasiou T, Harling L, Darzi A, Ashrafian H. Online social networking services in the management of patients with diabetes mellitus: systematic review and meta-analysis of randomised controlled trials. Diabetes Res Clin Pract 2014 Nov;106(2):200-211. [doi: 10.1016/j.diabres.2014.06.008] [Medline: 25043399]
21. Wu C, Wu Z, Yang L, Zhu W, Zhang M, Zhu Q, et al. Evaluation of the clinical outcomes of telehealth for managing diabetes: a PRISMA-compliant meta-analysis. Medicine (Baltimore) 2018 Oct;97(43):e12962 [FREE Full text] [doi: 10.1097/MD.0000000000012962] [Medline: 30412116]
22. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
23. Siriwardena LS, Wickramasinghe WA, Perera KL, Marasinghe RB, Katulanda P, Hewapathirana R. A review of telemedicine interventions in diabetes care. J Telemed Telecare 2012 Apr;18(3):164-168. [doi: 10.1258/jtt.2012.SFT110] [Medline: 22362832]
24. Tao D, Or CK. Effects of self-management health information technology on glycaemic control for patients with diabetes: a meta-analysis of randomized controlled trials. J Telemed Telecare 2013 Apr;19(3):133-143. [doi: 10.1177/1357633X13479701] [Medline: 23563018]
25. Baron J, McBain H, Newman S. The impact of mobile monitoring technologies on glycosylated hemoglobin in diabetes: a systematic review. J Diabetes Sci Technol 2012 Sep 1;6(5):1185-1196 [FREE Full text] [doi: 10.1177/193229681200600524] [Medline: 23063046]
26. de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of Internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014 Jan 16;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]
27. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
28. Suksomboon N, Poolsup N, Nge YL. Impact of phone call intervention on glycemic control in diabetes patients: a systematic review and meta-analysis of randomized, controlled trials. PLoS One 2014;9(2):e89207 [FREE Full text] [doi: 10.1371/journal.pone.0089207] [Medline: 24586596]
29. Lee JY, Lee SW. Telemedicine cost-effectiveness for diabetes management: a systematic review. Diabetes Technol Ther 2018 Jul;20(7):492-500. [doi: 10.1089/dia.2018.0098] [Medline: 29812965]
30. Walker CL, Kopp M, Binford RM, Bowers CJ. Home telehealth interventions for older adults with diabetes. Home Healthc Now 2017 Apr;35(4):202-210. [doi: 10.1097/NHH.0000000000000522] [Medline: 28353510]
31. Teljeur C, Moran PS, Walshe S, Smith SM, Cianci F, Murphy L, et al. Economic evaluation of chronic disease self-management for people with diabetes: a systematic review. Diabet Med 2017 Aug;34(8):1040-1049. [doi: 10.1111/dme.13281] [Medline: 27770591]
32. Macdonald EM, Perrin BM, Kingsley MI. Enablers and barriers to using two-way information technology in the management of adults with diabetes: a descriptive systematic review. J Telemed Telecare 2018 Jun;24(5):319-340. [doi: 10.1177/1357633X17699990] [Medline: 28347218]
33. Sood A, Watts SA, Johnson JK, Hirth S, Aron DC. Telemedicine consultation for patients with diabetes mellitus: a cluster randomised controlled trial. J Telemed Telecare 2018 Jul;24(6):385-391. [doi: 10.1177/1357633X17704346] [Medline: 28406066]

34. Kearns JW, Bowerman D, Kemmis K, Izquierdo RE, Wade M, Weinstock RS. Group diabetes education administered through telemedicine: tools used and lessons learned. Telemed J E Health 2012 Jun;18(5):347-353. [doi: 10.1089/tmj.2011.0165] [Medline: 22468984]
35. Fatehi F, Gray LC, Russell AW. Telemedicine for clinical management of diabetes - a process analysis of video consultations. J Telemed Telecare 2013 Oct;19(7):379-382. [doi: 10.1177/1357633X13506524] [Medline: 24218349]
36. Young H, Miyamoto S, Ward D, Dharmar M, Tang-Feldman Y, Berglund L. Sustained effects of a nurse coaching intervention via telehealth to improve health behavior change in diabetes. Telemed J E Health 2014 Sep;20(9):828-834 [FREE Full text] [doi: 10.1089/tmj.2013.0326] [Medline: 25061688]
37. Chen L, Chuang L, Chang C, Wang C, Wang I, Chung Y, et al. Evaluating self-management behaviors of diabetic patients in a telehealthcare program: longitudinal study over 18 months. J Med Internet Res 2013 Dec 9;15(12):e266 [FREE Full text] [doi: 10.2196/jmir.2699] [Medline: 24323283]
38. Fountoulakis S, Papanastasiou L, Gryparis A, Markou A, Piaditis G. Impact and duration effect of telemonitoring on HbA1c, BMI and cost in insulin-treated diabetes mellitus patients with inadequate glycemic control: a randomized controlled study. Hormones (Athens) 2015;14(4):632-643 [FREE Full text] [doi: 10.14310/horm.2002.1603] [Medline: 26188234]
39. Istepanian RS, Zitouni K, Harry D, Moutosammy N, Sungoor A, Tang B, et al. Evaluation of a mobile phone telemonitoring system for glycaemic control in patients with diabetes. J Telemed Telecare 2009;15(3):125-128. [doi: 10.1258/jtt.2009.003006] [Medline: 19364893]
40. Earle KA, Istepanian RS, Zitouni K, Sungoor A, Tang B. Mobile telemonitoring for achieving tighter targets of blood pressure control in patients with complicated diabetes: a pilot study. Diabetes Technol Ther 2010 Jul;12(7):575-579. [doi: 10.1089/dia.2009.0090] [Medline: 20597833]
41. Boaz M, Hellman K, Wainstein J. An automated telemedicine system improves patient-reported well-being. Diabetes Technol Ther 2009 Mar;11(3):181-186. [doi: 10.1089/dia.2008.0048] [Medline: 19216685]
42. Leichter SB, Bowman K, Adkins RA, Jelsovsky Z. Impact of remote management of diabetes via computer: the 360 study--a proof-of-concept randomized trial. Diabetes Technol Ther 2013 May;15(5):434-438. [doi: 10.1089/dia.2012.0323] [Medline: 23537419]
43. Codner E, Cerda T, Gaete X. Puberty in type 1 diabetes mellitus: advances in care are associated with changes in pubertal milestones and hormone profiles. Curr Opin Endocr Metab Res 2020 Oct;14:85-91. [doi: 10.1016/j.coemr.2020.06.002]
44. Diabetes Control Complications Trial Research Group, Nathan DM, Genuth S, Lachin J, Cleary P, Crofford O, et al. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993 Sep 30;329(14):977-986. [doi: 10.1056/NEJM199309303291401] [Medline: 8366922]
45. Viigimaa M, Sachinidis A, Toumpourleka M, Koutsampasopoulos K, Alliksoo S, Titma T. Macrovascular complications of type 2 diabetes mellitus. Curr Vasc Pharmacol 2020;18(2):110-116. [doi: 10.2174/1570161117666190405165151] [Medline: 30961498]
46. Mignerat M, Lapointe L, Vedel I. Using telecare for diabetic patients: a mixed systematic review. Health Policy Technol 2014 Jun;3(2):90-112. [doi: 10.1016/j.hlpt.2014.01.004]

Abbreviations

AMSTAR 2: A Measurement Tool to Assess Systematic Reviews
BP: blood pressure
CT: clinical trial
DFG: Deutsche Forschungsgemeinschaft
DM: diabetes mellitus
DRQoL: diabetes-related quality of life
EPHPP: Effective Public Health Practice Project
FBG: fasting blood glucose
GDM: gestational diabetes mellitus
HbA_{1c}: glycated hemoglobin A1c
HRQoL: health-related quality of life
ICER: incremental cost-effectiveness ratio
MA: meta-analysis
MAST: model for assessment of telemedicine
NICE: National Institute for Health and Care Excellence
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life
RCT: randomized controlled trial
SR: systematic review
T1DM: type 1 diabetes mellitus
T2DM: type 2 diabetes mellitus

WMD: weighted mean difference

Edited by G Eysenbach; submitted 05.08.20; peer-reviewed by KC Wong, K Yderstraede; comments to author 31.08.20; revised version received 16.10.20; accepted 28.10.20; published 19.02.21.

Please cite as:
Eberle C, Stichling S
Clinical Improvements by Telemedicine Interventions Managing Type 1 and Type 2 Diabetes: Systematic Meta-review
J Med Internet Res 2021;23(2):e23244
URL: http://www.jmir.org/2021/2/e23244/
doi:10.2196/23244
PMID:33605889

Original Paper

Efficacy of Smart Speaker–Based Metamemory Training in Older Adults: Case-Control Cohort Study

Jeongsim Kim[1*], MD; EunJi Shin[2*], BA; KyungHwa Han[2], MA; Soowon Park[3], PhD; Jung Hae Youn[2], PhD; Guixiang Jin[4], MSc; Jun-Young Lee[1], MD, PhD

[1]Department of Psychiatry, Seoul National University College of Medicine and SMG-SNU Boramae Medical Center, Seoul, Republic of Korea

[2]Department of Clinical Counseling Psychology, Cha University, Seoul, Republic of Korea

[3]Division of Teacher Education, College of Liberal Arts and Interdisciplinary Studies, Kyonggi University, Suwon, Republic of Korea

[4]Social Value Innovation Center, SK Telecom, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jun-Young Lee, MD, PhD
Department of Psychiatry
Seoul National University College of Medicine and SMG-SNU Boramae Medical Center
20 Boramae-Ro 5-Gil, Dongjak-gu
Seoul, 07061
Republic of Korea
Phone: 82 10 9612 0405
Fax: 82 2 6280 6099
Email: benji@snu.ac.kr

Abstract

Background: Metamemory training (MMT) is a useful training strategy for improving cognitive functioning in the older adult population. Despite the advantages, there are limitations imposed by location and time constraints.

Objective: This study aimed to develop a smart speaker–based MMT program and evaluate the efficacy of the program in older adults without cognitive impairment.

Methods: This study used a case-control cohort design. The smart speaker–based MMT program comprised 3 training sessions per day, 5 days a week, for 8 weeks. Each training session took approximately 15 minutes. This program was implemented using smart speakers, not human trainers. All participants completed the Mini-Mental State Examination, Subjective Memory Complaints Questionnaire, Verbal Learning Test, Digit Span Test, fluency tests, and a short-form version of the Geriatric Depression Scale before and after training.

Results: A total of 60 subjects (29 in the MMT group and 31 in the control group) participated in the study. The training group showed significant increases in the delayed free recall, digit span forward, digit span backward, and fluency test scores compared with the control group.

Conclusions: This study confirmed the efficacy of smart speaker–based MMT in older adults. Home-based smart speaker–based MMT is not limited with respect to location or constrained by space and may help older adults with subjective cognitive decline without requiring intervention by human professionals.

(J Med Internet Res 2021;23(2):e20177) doi:10.2196/20177

KEYWORDS

smart speaker; cognitive training; cognitive decline

Introduction

The gradual increase in the older adult population is leading to a growing problem with cognitive decline in the population. Many efforts have been made to address this concerning issue that is associated with older age.

The term “metamemory” was introduced by Flavell [1] and it refers to a type of metacognition, meaning the knowledge and awareness of an individual's own memory, including the contents and processes of their memory system [2]. Metamemory training (MMT) is a memory training program for the older population that is based on the metamemory concept, which consists of

metaknowledge, meta-monitoring, and meta-judgment [3]. In the meta-knowledge component, MMT participants obtain information on how cognitive aging affects their memory abilities and how the brain operates memorizing processes. Throughout this part of the training, older people are educated about efficient strategies for dealing with cognitive aging. In the program's meta-monitoring and meta-judgment components, participants develop the ability to monitor or judge their memory performance. Training programs help participants to understand memory and promote awareness of their own memory processing. This is accomplished through multimnemonic strategies (eg, storytelling, imagination, associations). MMT has been shown to have positive effects on everyday memory performance in older populations [4], as well as to improve memory and executive abilities in healthy older adults and adults with mild cognitive impairment [5-7]. MMT has also demonstrably enhanced the integrity of the white matter tract connecting the frontal and temporal cortices of the brain, which relate to the memory system [5].

Despite these advantages, there are some limitations to MMT. The first limitation involves the issues of accessibility and continuity. Many older adults have difficulty traveling to centers that run MMT programs. The second limitation relates to the insufficient number of experts available. Although the number of older adult populations experiencing cognitive decline continues to grow, the number of professionals nearby with expertise in the subject matter remains inadequate. It is difficult for older people to attend daily training sessions if they are living in rural areas where psychological experts are not readily available. Moreover, the active participation of people living in cities may be restricted by the usual business hours during which training is offered.

The emergence and integration of voice-user interfaces using artificial intelligence (AI) technology has led to the recent development of smart speakers. A smart speaker is a wireless and smart audio playback device that uses several types of connectivity for additional functions. They contain software that provides customized information or services to users while communicating with them by voice. Smart speakers have special features to enhance ease of use, connect to multiple types of audio sources, and provide additional functionality, and they are widely used in many countries. As international problems such as the COVID-19 situation have occurred, the "social distancing" culture has recently become the norm, and in such situations smart speakers that can be easily used at home and for various programs are becoming more and more popular [8].

Until now, there have been no smart speaker–based memory training programs, but there are some cognitive training tools that use computers or robots. Previous studies have reported that participants with high-functioning autism spectrum disorder who used computer-based virtual reality social training had improved social cognitive function [9,10]. In a previous study of participants with traumatic brain injury, improvements in memory abilities were reported following a virtual reality–based computer training program [11]. A recently published review of 11 studies on the effect of cognitive training with robots concluded that robot-based cognitive training in older adults with age-related cognitive decline could foster improved cognitive abilities [12]. None of these studies has overcome the above-mentioned limitations regarding location and time availability, such as requiring expert professional help or travel to a training center.

Although AI-based cognitive training has been found to be effective, the devices used for the training have some limitations. Users need to be able to use devices such as computers or robots, and many older adults are unfamiliar with their operation. These devices are expensive and challenging to deploy. On the other hand, smart speakers are affordable and offer a simple user interface that is verbally controlled. Therefore, older adults can use these devices with ease. Thus, smart speaker–based cognitive training may be a useful tool for overcoming the limitations of existing cognitive training programs.

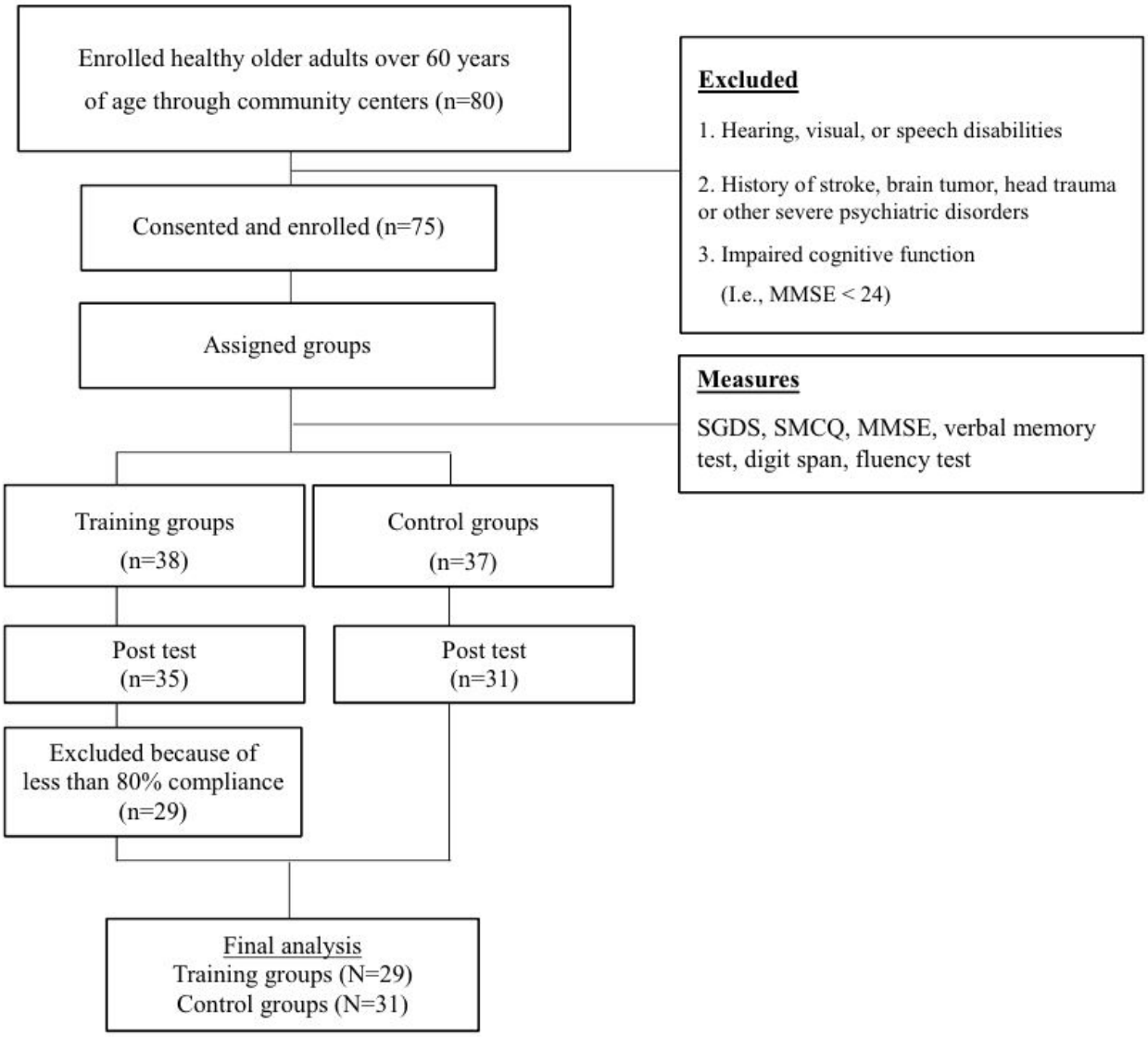
Therefore, the MMT provided by the smart speaker method should be effective in improving cognitive function, especially in the area of fluency, as indicated by the results of previous studies. On that basis, we developed a smart speaker–based MMT program, trained older adults by having them participate in the program for 8 weeks without in-person training, and tested the program's efficacy in a single-blind, case-control study.

Methods

Study Participants

Figure 1 shows the patient enrollment process. Eighty people over the age of 60 years with normal cognitive function were recruited through community advertising and enrolled through three community centers. During the recruitment process, 5 potential participants were excluded because they met one of the following exclusion criteria: (1) hearing, visual, or speech disabilities; (2) history of stroke, brain tumor, head trauma, or severe psychiatric disorders; or (3) impaired cognitive functioning (ie, Mini-Mental State Examination score <24). The remaining 75 participants were assigned into either the training group (n=38) or the control group (n=37) using the convenience sampling method. The training group assignees were provided with SK Telecom's smart speakers, while the control group assignees were not. Three participants in the training group and 6 participants in the control group were dropped from our analysis because they did not complete the poststudy evaluation. Thirty-five participants in the training group finished the smart speaker–based MMT and completed the pre/post cognitive tests. In the training group, we excluded 6 participants who had participation rates below 80%. Thirty-one participants in the control group completed the pre/post examinations. Finally, the analysis was conducted on 29 participants in the training group and 31 participants in the control group using per protocol analysis.

Figure 1. Study enrollment process. MMSE: Mini-Mental State Examination; SGDS: Short-Form Geriatric Depression Scale; SMCQ: Subjective Memory Complaints Questionnaire.



Smart Speakers

The emergence and integration of voice interfaces using AI technology has led to the recent development of smart speakers. Smart speakers contain software that provides customized information or services to users while communicating with them by voice. The devices receive requests from users in the form of voice, text, or other communication forms and then performs the requested tasks.

As AI technology is at the core of smart speakers, the spoken language dialogue system (SLDS) proposed by Bertrand et al [13] is most similar to the system used in this study. SLDS functions include (1) speech recognition (including language understanding), (2) speech synthesis, and (3) dialogue management; all are similar to the current smart speaker system structure. Speech recognition technology refers to technology that receives human speech and converts it into text. This type of human–machine interaction technology can control various devices and services using human speech [14]. Speech synthesis mechanically produces human speech using a computer and is an active research focus due to the development of automation technology using deep neural networks. Recent synthesis technology that enables speech to reflect a specific tone has been developed to a level that makes various types of technical utilization possible, such as emotional expression [15]. Dialogue management receives text sentences from recognized human speech, understands the speaker's intentions, and responds appropriately. Goal-oriented dialogue processing technology, which understands speech and maintains conversation on a related topic to accomplish a specific purpose or task, is used to implement smart speaker dialogue.

Based on voice recognition technology, smart speakers have voice interactions that can extract meaning from human voices, which leads to understanding between humans and machines. We call this a voice-user interface [16]. This voice interaction consists of voice recognition, voice synthesis, and dialogue management technologies. A system that provides voice interaction based on these technologies is called a conversational agent. Beyond their use as an interface, voice interactions with smart speakers are the key interaction determining the user's main service experience. From the user's perspective, this is the most natural method of human interaction with machines.

NUGU: The Smart Speaker Platform

This study was conducted using NUGU (SK Telecom [17]), a smart speaker platform that is widely used in Korea. According to a survey by SK telecom, there are currently about 1 million consumers with NUGU speakers. In August 2019, 670 million individuals used a NUGU speaker at least once a month (monthly active users [MAUs]), and in the first half of 2020, there were an average of 6 million MAUs of NUGU [18]. NUGU analyzes users' natural language (voice or text) requests, interprets them, and then provides users with information or related services. For detailed information, see Multimedia Appendix 1.

Brain Toktok: The Smart Speaker–Based MMT Program

In this study, the smart speaker–based MMT is called "Brain Toktok" or "Brain Opal." It was commercialized on July 30, 2020. Users of NUGU speakers can subscribe to it [19]. Brain Toktok is based on a multistrategy MMT program developed by Youn and colleagues [7]. Previous offline MMT studies were conducted by trained psychological experts and MMT has already been well-validated [5-7]. The offline MMT program applied several memory strategies based on the metamemory concept. The program consists of 10 weekly sessions. Each session is comprised of three parts, and it takes 90 minutes to complete a session. The goal of the first part of each session is to strengthen meta-knowledge. The second and third parts of each session focus on meta-monitoring, meta-judgment, and memory training via the practice of external and internal strategies. Brain Toktok was developed through a series of analytic stages (content, learner, technology, and environment analysis) and design stages (information, interaction, synchronous, and evaluation design), followed by expert consultation. In this process, the goal of Brain Toktok was to focus on the steady participation of users in training rather than to increase the user's correctness rate. Finally, 11 cognitive training exercises were developed. Three programs are randomly selected for use among the 11 cognitive training exercises possible per session. The provided training program was selected randomly without a fixed standard, and the content provided for each session was not recorded separately. Users can respond to multiple-choice or open-response questions specific to the programs provided and consist of more than 100 questions per program. For detailed information about the individual programs, see Multimedia Appendix 2.

Measures

Short-Form Geriatric Depression Scale (SGDS)

Many studies have shown a high association between depression and cognitive function, and depression can independently affect cognitive function. The Geriatric Depression Scale (GDS) [20] is one of the most widely used instruments worldwide for screening late-life depression. This scale involves a 30-item easy-to-administer inventory; it has been widely used in both communities and institutions [21,22]. To simplify this screening device for depression, a short-form version was also developed (SGDS), which was extracted from the original GDS. There have been many studies proving that the SGDS is an adequate substitute for the 30-item GDS [23-26]. This scale was standardized for Korean older adults [27]. The SGDS consists of 15 questions, and all questions can be answered with yes or no. A score of 10 or above indicates major depression. Item analysis also confirmed consistency. All 15 items of SGDS were significantly correlated with the full GDS, and its sensitivity (91%) and specificity (82%) were very high [28]. Therefore, the SGDS was used in this study to assess depressed mood in older adults and to exclude patients with major depressive disorder.

Subjective Memory Complaints Questionnaire (SMCQ)

The SMCQ is a brief, self-rated questionnaire for assessing subjective memory complaints (SMC), including memory problems in general and daily living. It consists of 14 items reflecting different aspects of SMC, which represents the metacognition of general and specific memories. Four items assess subjective judgment of memory impairment and 10 items assess memory deficit in everyday life. Higher scores indicate a greater perceived cognitive decline. Participants with an SMCQ score of 6 or above were assigned to the SMC group. The SMCQ has been validated and adapted for the Korean population [29]. Consequently, SMCQ is a reliable and important tool for SMC evaluation by measuring subjective cognitive problems. Therefore, we used the SMCQ to measure participants' SMC at baseline to try to provide homogeneity between the training and control groups. In addition, SMCQ is meaningful in that it can reflect the degree of metamemory because it is measured by the recognition and judgment of its memory function.

Mini-Mental State Examination (MMSE)

The MMSE was designed to assess cognitive functioning [30]. The Korean version of the MMSE was developed and validated in older Korean adults by Lee et al [31]. The score ranges from 0 to 30, with higher scores indicating better cognition. The MMSE score reflects functioning in six areas, each scored separately: orientation (10 points), short-term memory registration (3 points), memory recall (3 points), attention and calculation (5 points), language (8 points), and copying a double pentagon (1 point).

Verbal Learning Test: Rappel Indicé (RI)

The RI was used to assess verbal learning and memory. The RI was developed by Adam et al [32] and was validated for use with older Korean adults [33]. The RI includes 24 items across six different categories for recall and uses category-cued recall. For example, the fruit category consists of four words: grape, banana, watermelon, and oriental melon. The participants first complete a learning phase in which four items are presented on a slide. Of these, they are asked to name the item included in the semantic category provided by the examiner. For example, for the item "grape," the examiner would ask, "Which item is in the category of fruit?" Correct answers are then scored, with total scores ranging from 0 to 24. Approximately 20 minutes later, participants are asked to recall the four items from each category in any order. There is no specific time limit, and the examiner proceeds to the next category if the participant is unable to list the remaining items from the category. Correctly

recalled answers are scored from 0 to 24. In this study, verbal learning and delayed free recall were assessed using the RI.

Working Memory: Digit Span Test

The Digit Span Test is a component of the Elderly Memory Disorder Scale, which has been developed and validated to measure memory and cognitive functions in older Korean adults [34]. The test consists of a forward and backward digit order recall. The examiner calls number digits and asks participants to repeat the list both forward and backward. Scores range from 0 to 14.

Executive Function: Fluency Test

The fluency test also comes from the Elderly Memory Disorder Scale [34]. This test evaluates the ability to form and fluently utter words compatible with given criteria. The test consists of three parts. The first section involves listing as many words as possible belonging to a given semantic category in 30 seconds (usually, this is made up of the names of objects). In the second section, the objective is to list as many words as possible belonging to a given phonetic category (ie, words containing a given sound) in 30 seconds. In the third section, the objective is to list as many words associated with a given item as possible in 30 seconds. In this study, the cues were the names of fruits, words containing the Korean letter “Ma,” and words associated with the word “fox.” The score for this test is calculated as the number of correct words listed for each category and item.

Procedures

This study was conducted from October 2019 to March 2020. Individuals who agreed to participate were assigned into two groups using convenience sampling. The control group did not receive any training; neuropsychological measures were simultaneously taken in the control group and in the training group before and after the training was completed in the latter group. As shown in Figure 1, 6 control group participants were excluded from the study because they did not complete the postevaluation. At the start of the study, participants in the training group were instructed on the use of the smart speaker. Neuropsychological measures (ie, SGDS, SMCQ, MMSE, RI, Digit Span Test, and fluency test) were assessed twice during the study: before the training (pretraining evaluation) and after the training (posttraining evaluation).

The training group received motivation reinforcement education training at the beginning of the smart speaker–based MMT. Group education took place for 1 hour and 30 minutes, and lectures were conducted on the definition, purpose, effectiveness, and necessity of metacognitive training. Afterward, participants had an opportunity to provide feedback. The 8-week training program was designed to include training sessions 3 times a day, 5 days a week (weekdays), for approximately 15 minutes. The patient flexibly set the time without setting a specific time for program participation. During Brain Toktok, the program may be suspended if there are external factors such as personal circumstances of the user or instability of the internet connection. If the program is interrupted due to external factors, it starts again from the beginning when participation resumes. To check the training group’s compliance, the number and duration of times each patient turned on the program and the number of times they completed the session was stored in the server. Participants who did not complete a training session in a given day received an encouraging program message. In the control group, the monitoring system was not implemented as in the experimental group, but the control group was registered in the local welfare center and the situation of the control group was continuously confirmed by the welfare center staff.

All participants provided written consent before study participation. The study was conducted following the Declaration of Helsinki and approved by the ethics review board of the Seoul National University Boramae Medical Center. Participants received a gift card worth US $85 as compensation for their participation in the study.

Focus Group Interviews

Focus group interviews were conducted with participants in the experimental group to assess their experience with and the effectiveness of the smart speaker–based MMT program. After completing the smart speaker–based MMT course, interviews were conducted with 28 participants who had agreed to the interview during the course. It was composed of eight open-ended questions, and for each question, the subject was asked to freely describe their thoughts about the item. An experienced clinical psychologist conducted the interviews and recorded both their observations and the participants’ responses.

Statistical Analysis

Statistical analyses were conducted using SPSS software (version 20.0; IBM Corp). Differences between groups in demographics and neuropsychological measures (ie, age, education, gender, and MMSE score) were assessed using independent *t* tests or χ^2 tests. Repeated-measures analysis of variance (rmANOVA) was used to evaluate differences between the pre-and posttraining neuropsychological scores in the training and control groups. For model I, no adjusted factors were used. For model II, adjusted factors were used for SGDS and SMCQ scores. Because the SGDS and SMCQ scores varied between the groups at baseline, adjusting for the effects of SGDS and SMCQ scores allowed us to see the true effect of MMT on the memory test score. Statistical tests were two-tailed, with $P<.05$ indicating significant results.

Results

Participant Characteristics

Figure 1 summarizes the characteristics of the study participants. No difference was found in the drop-out rate between the two groups (MMT group: 9 dropouts, control group: 6 dropouts; $\chi^2=0.65$; $P=0.42$). All but 6 participants in the MMT group achieved more than 80% compliance with the attendance requirement. The 6 subjects who failed to achieve more than 80% compliance did so because of personal reasons (ie, hospital admission, sickness, preparation for moving, etc). The details of compliance are shown in Multimedia Appendix 2. Table 1 summarizes the demographic and clinical characteristics of the MMT and control groups. There were no differences between the groups regarding gender distribution, age, education, or

MMSE scores. There was a difference in the SMCQ and SGDS scores between the groups. Therefore, we adjusted the SMCQ and SGDS scores in regression model II.

Table 1. Demographic characteristics of the metamemory training (MMT) and control groups.

Characteristics	Total (N=60)	Groups		Mean difference (95% CI) or χ^2	*P* value
		MMT (n=29)	Control (n=31)		
Female gender, n (%)	50 (83.3)	24 (82.8)	26 (83.9)	0.01	.91
Age, years (SD)	71.21 (5.72)	70.48 (6.08)	71.94 (5.36)	1.453 (–1.503 to 4.409)	.33
Education, years (SD)	10.60 (3.83)	10.17 (4.17)	11.03 (3.49)	0.860 (–1.121 to 2.841)	.39
SGDS[a] score, mean (SD)	2.26 (3.37)	3.21 (4.16)	1.32 (2.59)	–1.884 (–3.662 to –0.106)	.04
SMCQ[b] score, mean (SD)	17.84 (2.64)	18.69 (3.13)	17.00 (2.16)	–1.690 (–3.072 to –0.308)	.02
MMSE[c] score, mean (SD)	27.90 (1.50)	27.76 (1.66)	28.03 (1.33)	0.274 (–0.501 to 1.049)	.48

[a]SGDS: Short-Form Geriatric Depression Scale.

[b]SMCQ: Subjective Memory Complaints Questionnaire.

[c]MMSE: Mini-Mental State Examination.

Effects on Improving Cognitive Function

As shown in Table 2, we found significant interactions between group and score change for the tests of delayed free recall, fluency, digit span forward, and digit span backward. Because SMC may affect cognitive performance regardless of memory training, even after an improvement of depressive symptoms, the change in the SGDS score was controlled for by treating it as a covariate. After controlling for the effects of baseline SMCQ and SGDS scores, the interactions remained significant (model II in Table 3). The change in verbal immediate recall score did not interact significantly with the group.

The scores for delayed free recall, fluency, digit span forward, and digit span backward were significantly higher in the group that participated in the MMT program for 8 weeks; there were no significant cognitive changes in the control group after 8 weeks (Tables 2 and 3, and Figure 2).

Table 2. Changes in cognitive function in the metamemory training (MMT) and control groups.

Cognitive function (measures)	MMT group (n=29)			Paired *t* test		Control group (n=31)			Paired *t* test	
	Pre, mean (SD)	Post, mean (SD)	MD[a] (95% CI)	MMT group (*d*[b] 1)	Control group (*P* value 1)	Pre, mean (SD)	Post, mean (SD)	MD (95% CI)	MMT group (*d* 2)	Control group (*P* value 2)
Verbal memory										
Learning	22.03 (4.49)	22.00 (2.45)	0.03 (–1.87 to 1.94)	.01	.97	23.23 (1.43)	23.45 (1.18)	–0.23 (–0.89 to 0.44)	.12	.52
Delayed cued recall	12.21 (4.47)	15.34 (3.71)	–3.14 (–5.3 to –0.98)	.84	<.001	13.00 (3.92)	13.35 (3.73)	–0.36 (–2.3 to 1.59)	.09	.62
Attention/executive functions										
Digit span forward	5.97 (1.64)	7.24 (2.26)	–1.28 (–2.32 to –0.24)	.50	.01	6.65 (1.58)	6.68 (1.11)	–0.03 (–0.73 to 0.66)	.02	.90
Digit span backward	3.41 (1.38)	4.55 (1.38)	–1.14 (–1.86 to –0.41)	.84	<.001	3.70 (1.32)	3.84 (1.34)	–0.1 (–0.77 to 0.58)	.08	.69
Language										
Fluency	15.76 (3.90)	18.86 (4.35)	–3.10 (–5.28 to –0.93)	.85	<.001	18.03 (4.50)	17.55 (5.00)	0.48 (–1.93 to 2.90)	.12	.51

[a]MD: mean difference.

[b]Cohen *d*.

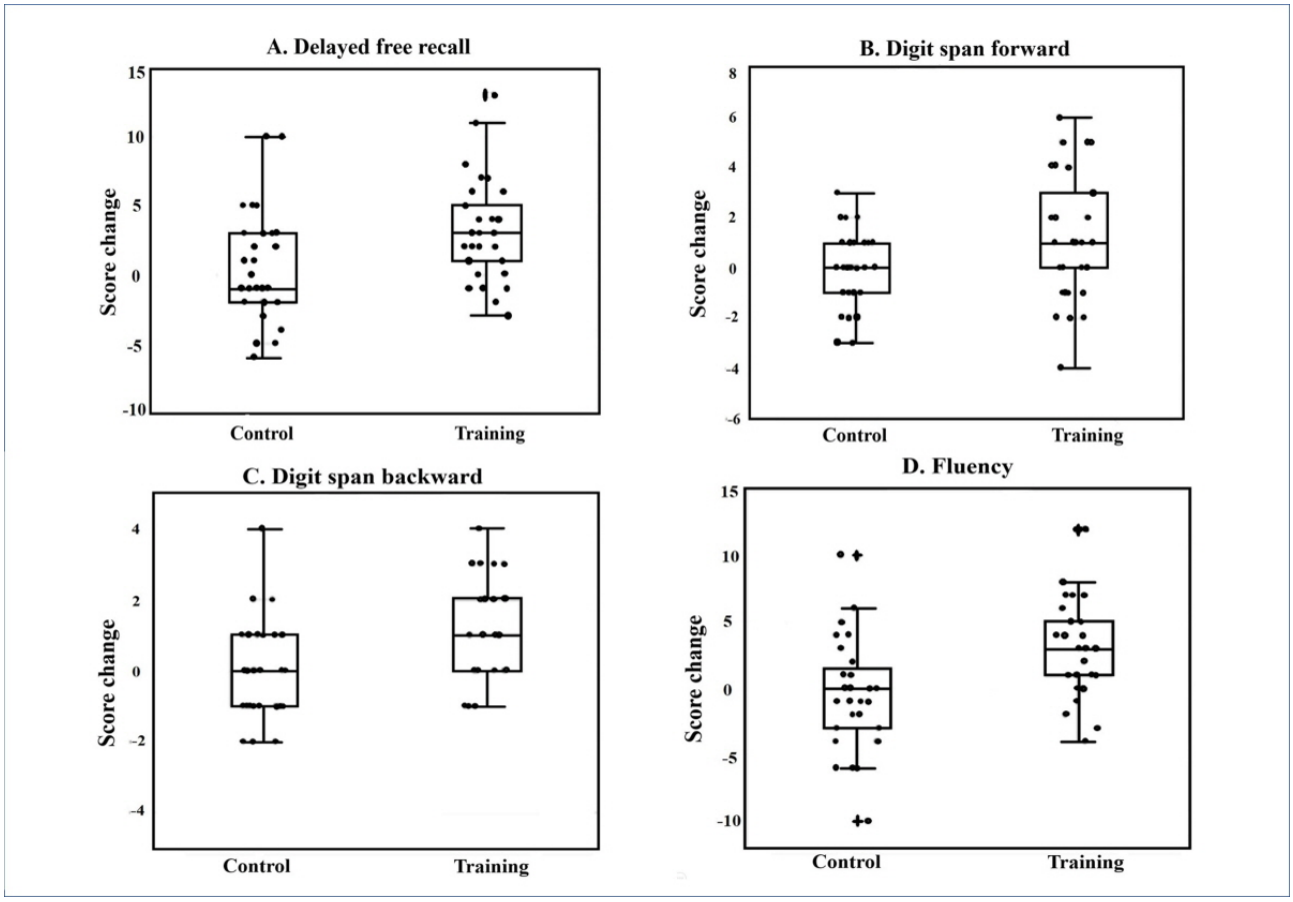
Table 3. Interaction between cognitive changes and study groups.

Measures of cognitive function	Main effect of cognitive change/interaction between cognitive change and study group					
	Model I[a]			Model II[b]		
	F	η^2	*P* value	*F*	η^2	*P* value
Verbal memory						
Learning	0.03/0.06	0.001/0.001	.86/.81	0.13/0.19	0.002/0.003	.72/.67
Delayed free recall	12.38/7.86	0.18/0.12	.001/.007	0.22/8.64	0.004/0.13	.64/.005
Attention/ executive functions						
Digit span forward	6.08/5.49	0.10/0.09	.02/.02	1.75/4.15	0.03/0.07	.19/.046
Digit span backward	12.72/9.04	0.18/0.14	.001/.004	0.42/6.19	0.007/0.10	.52/.02
Language						
Fluency	6.90/12.94	0.11/0.18	.01/.001	0.75/9.24	0.01/0.14	.39/.004

[a]Model I: no adjusted factors.

[b]Model II: adjusted for Short-Form Geriatric Depression Scale and Subjective Memory Complaints Questionnaire scores.

Figure 2. Box and scatter plots of cognitive function changes in metamemory training and control groups during the 8 weeks of the study. (A) Delayed free recall (*P*<.001; *t*=0.50; Cohen *d*=0.84). (B) Digit span forward (*P*=.01; *t*=2.71; Cohen *d*=0.50). (C) Digit span backward (*P*<.001; *t*=4.52; Cohen *d*=0.84). (D) Fluency (*P*<.001; *t*=4.58; Cohen *d*=0.85).



Focus Group Interviews

Users of smart speakers reported that they were easier to use and less burdensome than other information and communications technology tools such as computers and mobile phones. For example, one user reported, “The smart speaker was convenient because I did not need to press the keyboard or the button, and usually I didn’t use the cell phone very well. Rather, I used it more frequently than the cell phone.”

Although the smart speaker is a machine, users reported that communicating with it was like a human-to-human conversation: “Television is what you see, and you can't talk but the smart speaker is able to communicate.” Some users preferred to communicate with the smart speaker rather than real people,

and they enjoyed it: “If I get it wrong repeatedly, people get easily tired. The AI speaker keeps listening and encouraging;” “Smart speakers feel like comfortable friends, sometimes like a family;” and “I don't like to communicate to people very much, but I enjoyed communicating to the smart speaker when I needed it.”

Discussion

Principal Results

According to the study results, our smart speaker–based MMT program improved memory ability, executive ability, and working memory. After adjusting for SMCQ and SGDS scores, the effects were still significant. These results indicate that cognitive improvements may be achieved through the effects of pure cognitive training. Furthermore, smart speaker–based MMT, which can easily be carried out without trained experts, fostered high compliance. Hence, it offers a good cognitive training tool without the typical limitations imposed by location or time constraints. To the best of our knowledge, this is the first report on a cognitive training program using a smart speaker. We hope that this study offers a useful guideline for future research on cognitive training programs based on smart speakers.

Comparison with Prior Research

Previous studies on the effectiveness of smart speaker–based MMT reported that healthy older adults achieved improvements in memory ability, verbal fluency performance, and working memory [7]. The results of this study are similar to those of previous studies. Users of smart speaker–based MMT may still experience the same cognitive improvements as people who undertake in-person courses conducted by experts.

Effects on Delayed Free Recall and Verbal Memory

In this study, the scores for delayed free recall and verbal memory were significantly higher in this healthy older adult population after the training. Decreased delayed recall scores are considered the best predictor of progression from a dementia-free period to Alzheimer disease dementia (ADD) [35]. A previous study reported that expert MMT was associated with significantly improved cognition in subjects with mild cognitive impairment (MCI) [6]. These results suggested that smart speaker–based MMT may slow not only cognitive decline in healthy individuals but the progression of MCI to ADD.

Effects on Verbal Fluency

Among the improvements in executive functions, the improvement in verbal fluency was significant. Previous studies have reported that verbal fluency was significantly associated with social activity [36]. According to the US Census Bureau, in 2010, 30% of people aged 65 years and older lived alone at the time of the census. As people get older, their likelihood of living alone increases [37]. Due to the social conditions associated with aging, older people are more likely to be vulnerable to verbal fluency problems. Smart speaker–based MMT can effectively support the vulnerable cognitive functioning of older adults who live alone. Also, older people who live alone and repeat relatively simple life patterns have a higher risk of depression and poor cognitive abilities compared with older people who do not-live alone. Smart speaker–based MMT is conducted through conversations with users, which can serve as a new stimulus for the elderly user who lives alone [38]. According to our focus group interviews, smart speaker users reported that communicating with it was like conversing with another person, so our smart speaker–based MMT may help those who live alone.

Effects on Digit Span Forward and Backward

Digit span forward and backward tests are the most commonly used tests in clinical neuropsychology to assess working memory capacity [39]. Working memory is a key focus of memory improvement strategies. It can be trained through smart speaker–based MMT and then employed to increase cognitive reserve to prevent dementia. Furthermore, working memory is one of the core mechanisms involved in higher-order cognitive abilities, such as problem-solving, fluid intelligence, and reading comprehension [40,41]. It is also one of the cognitive processes that suffers a clear and linear decline with aging [42,43]. In this study, digit span forward and backward tests were significantly improved in the group that participated in 8 weeks of training. Smart speaker–based MMT may help maintain or improve higher-order cognitive abilities by enhancing working memory capacity.

Limitations

There are some limitations to this study, the first being the small sample size. However, this is the first time a cognitive training program has been developed using the smart speaker, so this paper provides meaningful guidelines for future studies. Second, this study did not adopt an active control design, and the control group took no action. Further studies should use an active control design. Third, the SMCQ and SGDS scores were not homogenous between the groups. To remove the effects of SMCQ and SGDS scores, the SMCQ and SGDS values were adjusted during the analysis. We found that the efficacy of the MMT program using the smart speaker was not affected by the SMCQ and SGDS scores. Therefore, the results of this study are reliable.

Fourth, neuropsychological tests performed before and after the experiment are not a direct measure of metamemory function. Still, we used neuropsychological tests to assess the functioning of general memory areas. Specifically, the test of cued recall is the most effective test of memory function status. We also employed the cued recall test to discover the effectiveness of smart speaker–based-MMT. Although a direct measure of metamemory function was not used, the purpose of this study was to examine the enhancement of memory function through MMT. We consider the study methodology to be suitable for its purpose. Finally, we assigned participants to the two study groups by convenience sampling. We are aware of one other recent study that has addressed this issue [44]. Although cross-sectional in design, the previous study found few significant differences when contrasting the demographic characteristics and psychological performance of older adults in a random sample instead of two convenience samples. Specifically, the previous study observed no differences in word list recall results between the random and convenience samples

recruited for a study of memory and aging. In further studies, to increase the power and proof of the effectiveness of smart speaker–based MMT, a large sample size and a more elaborate research design such as a randomized controlled study with an active control design should be implemented.

Conclusion

Smart speaker–based MMT, without location requirements and time constraints, may address older adult memory problems and possibly improve their quality of life by helping them cope with the cognitive issues associated with aging. Our smart speaker–based MMT may also be useful in delaying the onset of dementia.

Our study's strength is that it is the first report on a cognitive training program using a smart speaker. In the future, the use of smart speakers will be more commercialized, expanding with the paradigm shift to "social distancing." The problem of memory deterioration due to the increase in the elderly population will become more critical. In the future, the combination of a smart speaker and the memory training program will be of great help to older adults. We hope that this study offers a useful guideline for future studies on cognitive training programs based on smart speakers.

Acknowledgments

This study was supported by SK Telecom, Republic of Korea.

Conflicts of Interest

None declared.

Multimedia Appendix 1
NUGU: Smart speaker platform.
[DOCX File , 1246 KB - jmir_v23i2e20177_app1.docx]

Multimedia Appendix 2
Supplementary table.
[DOCX File , 32 KB - jmir_v23i2e20177_app2.docx]

References

1. Flavell JH. Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. American Psychologist 1979;34(10):906-911. [doi: 10.1037/0003-066x.34.10.906]
2. Hertzog C, Dixon RA, Hultsch DF. Relationships between metamemory, memory predictions, and memory task performance in adults. Psychol Aging 1990 Jun;5(2):215-227. [doi: 10.1037//0882-7974.5.2.215] [Medline: 2378687]
3. Gilleen J, David A, Greenwood K. Self-reflection and set-shifting mediate awareness in cognitively preserved schizophrenia patients. Cogn Neuropsychiatry 2016 May;21(3):185-196. [doi: 10.1080/13546805.2016.1167031] [Medline: 27112316]
4. McDougall GJ, Kang J. Memory Self-Efficacy and Memory Performance in Older Males. Int J Mens Health 2003 May 1;2(2):131-147 [FREE Full text] [doi: 10.3149/jmh.0202.131] [Medline: 19043600]
5. Youn J, Ryu S, Lee J, Park S, Cho S, Kwon H, et al. Brain structural changes after multi-strategic metamemory training in older adults with subjective memory complaints: A randomized controlled trial. Brain Behav 2019 May;9(5):e01278 [FREE Full text] [doi: 10.1002/brb3.1278] [Medline: 30916450]
6. Youn J, Park S, Lee J, Cho S, Kim J, Ryu S. Cognitive Improvement in Older Adults with Mild Cognitive Impairment: Evidence from a Multi-Strategic Metamemory Training. J Clin Med 2020 Jan 28;9(2):362 [FREE Full text] [doi: 10.3390/jcm9020362] [Medline: 32013035]
7. Youn J, Lee J, Kim S, Ryu S. Multistrategic memory training with the metamemory concept in healthy older adults. Psychiatry Investig 2011 Dec;8(4):354-361 [FREE Full text] [doi: 10.4306/pi.2011.8.4.354] [Medline: 22216046]
8. Sarah P. COVID-19 quarantine boosts smart speaker usage among U.S. adults, particularly younger users. techcrunch. URL: https://techcrunch.com/2020/04/30/covid-19-quarantine-boosts-smart-speaker-usage-among-u-s-adults-particularly-younger-users/ [accessed 2020-05-01]
9. Kandalaft MR, Didehbani N, Krawczyk DC, Allen TT, Chapman SB. Virtual reality social cognition training for young adults with high-functioning autism. J Autism Dev Disord 2013 Jan;43(1):34-44 [FREE Full text] [doi: 10.1007/s10803-012-1544-6] [Medline: 22570145]
10. Yang YJD, Allen T, Abdullahi SM, Pelphrey KA, Volkmar FR, Chapman SB. Neural mechanisms of behavioral change in young adults with high-functioning autism receiving virtual reality social cognition training: A pilot study. Autism Res 2018 May;11(5):713-725 [FREE Full text] [doi: 10.1002/aur.1941] [Medline: 29517857]
11. Man DWK, Poon WS, Lam C. The effectiveness of artificial intelligent 3-D virtual reality vocational problem-solving training in enhancing employment opportunities for people with traumatic brain injury. Brain Inj 2013;27(9):1016-1025. [doi: 10.3109/02699052.2013.794969] [Medline: 23662639]

12. Vogan AA, Alnajjar F, Gochoo M, Khalid S. Robots, AI, and Cognitive Training in an Era of Mass Age-Related Cognitive Decline: A Systematic Review. IEEE Access 2020;8:18284-18304. [doi: 10.1109/access.2020.2966819]
13. Bertrand G, Nothdurft F, Minker W. What do you want to do next? Providing the user with more freedom in adaptive spoken dialogue systems. 2012 Aug 06 Presented at: Proc - 8th Int Conf Intell Environ IE; 2012; Guanajuato p. 290-296.
14. Stolcke A, Droppo J. Comparing human and machine errors in conversational speech transcription. 2017 Aug 24 Presented at: Proc Annu Conf Int Speech Commun Assoc INTERSPEECH; 2017; Stockholm p. 137-141.
15. Oord AVD, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, et al. WaveNet: A Generative Model for Raw Audio Based on PixelCNN Architecture. arXiv 2016 2016 Sep 19:13-15.
16. Schnelle D, Lyardet F. Voice user interface design patterns. 2006 Jul 09 Presented at: Eur th Eur Conf Pattern Lang Programs; 2006; Irsee p. 2006-2011. [doi: 10.1145/1415472.1415511]
17. SK Telecom Group. SKT NUGU. SK telecom Co. URL: https://www.nugu.co.kr/ [accessed 2016-09-01]
18. Hayoung K. The 3rd anniversary of the launch of SKT “NUGU”... “6.7 million users more than once a month.”. Digital Daily. 2019 Sep 05. URL: http://www.ddaily.co.kr/news/article/?no=185491 [accessed 2019-09-05]
19. Huegang S. SKT, Lanched “NUGU opal”, AI service for older adults. Newdaily. URL: http://biz.newdaily.co.kr/site/data/html/2020/07/30/2020073000131.html [accessed 2020-07-30]
20. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4] [Medline: 7183759]
21. Hyer L, Blount J. Concurrent and discriminant validities of the geriatric depression scale with older psychiatric inpatients. Psychol Rep 1984 Apr;54(2):611-616. [doi: 10.2466/pr0.1984.54.2.611] [Medline: 6739655]
22. Zgourides G, Spofford M, Doppelt L. The Geriatric Depression Scale: discriminant validity and elderly day-treatment clients. Psychol Rep 1989 Jun;64(3 Pt 2):1082. [doi: 10.2466/pr0.1989.64.3c.1082] [Medline: 2762456]
23. Sherry A G. The geriatric depression scale (GDS) validation of a geriatric depression screening scale: A preliminary report. Best Pract Nurs Care to Older Adults 2019(4):1-3. [doi: 10.1016/b978-0-323-03671-9.50021-6]
24. Lesher EL, Berryhill JS. Validation of the geriatric depression scale-short form among inpatients. J. Clin. Psychol 1994 Mar;50(2):256-260. [doi: 10.1002/1097-4679(199403)50:2<256::aid-jclp2270500218>3.0.co;2-e]
25. Fountoulakis KN, Tsolaki M, Iacovides A, Yesavage J, O'Hara R, Kazis A, et al. The validation of the short form of the Geriatric Depression Scale (GDS) in Greece. Aging (Milano) 1999 Dec;11(6):367-372. [doi: 10.1007/BF03339814] [Medline: 10738851]
26. Burke WJ, Roccaforte WH, Wengel SP. The short form of the Geriatric Depression Scale: a comparison with the 30-item form. J Geriatr Psychiatry Neurol 1991;4(3):173-178. [Medline: 1953971]
27. Bae JN, Cho MJ. Development of the Korean version of the Geriatric Depression Scale and its short form among elderly psychiatric patients. J Psychosom Res 2004 Sep;57(3):297-305. [doi: 10.1016/j.jpsychores.2004.01.004] [Medline: 15507257]
28. Cho M, Bae J, Suh G, Hahm B, Kim J, Lee D, et al. Validation of Geriatric Depression Scale, Korean Version(GDS) in the Assessment of DSM-III-R Major Depression. J Korean Neuropsychiatr Assoc 1999 Jan 30;38(1):48-63.
29. Youn JC, Kim KW, Lee DY, Jhoo JH, Lee SB, Park JH, et al. Development of the Subjective Memory Complaints Questionnaire. Dement Geriatr Cogn Disord 2009;27(4):310-317. [doi: 10.1159/000205512] [Medline: 19252402]
30. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
31. Lee DY, Lee KU, Lee JH, Jhoo JH, Kim KW, Youn JC, et al. A Normative Study of the Mini-Mental State Examination in the Korean Elderly. J Korean Neuropsychiatr Assoc 2002 May 01;41:508-525.
32. Adam S, Van der Linden M, Ivanoiu A, Juillerat A, Bechet S, Salmon E. Optimization of encoding specificity for the diagnosis of early AD: the RI-48 task. J Clin Exp Neuropsychol 2007 Jul 12;29(5):477-487. [doi: 10.1080/13803390600775339] [Medline: 17564913]
33. Park S, Kim I, Park HG, Shin SA, Cho Y, Youn J, et al. Development and Validation of the Rappel Indicé-24: Behavioral and Brain Morphological Evidence. J Geriatr Psychiatry Neurol 2016 May 08;29(3):160-168. [doi: 10.1177/0891988716632912] [Medline: 26956224]
34. Choi JY. Elderly memory disorder scale. Seoul: Hakjisa; Jan 10, 2007.
35. Elias MF, Beiser A, Wolf PA, Au R, White RF, D'Agostino RB. The preclinical phase of alzheimer disease: A 22-year prospective study of the Framingham Cohort. Arch Neurol 2000 Jun;57(6):808-813. [doi: 10.1001/archneur.57.6.808] [Medline: 10867777]
36. Kelly ME, Duff H, Kelly S, McHugh Power JE, Brennan S, Lawlor BA, et al. The impact of social activities, social networks, social support and social relationships on the cognitive functioning of healthy older adults: a systematic review. Syst Rev 2017 Dec 19;6(1):259 [FREE Full text] [doi: 10.1186/s13643-017-0632-2] [Medline: 29258596]
37. Census Bureau United States. United states Census Bureau in 2010 Internet. United States Census Bureau. URL: https://www.census.gov/ [accessed 2010-12-31]
38. Jung Y, Kim J. [Comparison of cognitive levels, nutritional status, depression in the elderly according to living situations]. Taehan Kanho Hakhoe Chi 2004 Jun;34(3):495-503. [doi: 10.4040/jkan.2004.34.3.495] [Medline: 15314306]

39. Hilbert S, Nakagawa TT, Puci P, Zech A, Bühner M. The Digit Span Backwards Task. European Journal of Psychological Assessment 2015 Jul;31(3):174-180. [doi: 10.1027/1015-5759/a000223]
40. Borella E, Ghisletta P, de Ribaupierre A. Age differences in text processing: the role of working memory, inhibition, and processing speed. J Gerontol B Psychol Sci Soc Sci 2011 May;66(3):311-320. [doi: 10.1093/geronb/gbr002] [Medline: 21339301]
41. Graf HP, Ohta N, editors. Lifespan Development of Human Memory. Cambridge, Massachusetts: The MIT Press; 2002.
42. Borella E, Carretti B, De Beni R. Working memory and inhibition across the adult life-span. Acta Psychol (Amst) 2008 May;128(1):33-44. [doi: 10.1016/j.actpsy.2007.09.008] [Medline: 17983608]
43. Mammarella IC, Borella E, Pastore M, Pazzaglia F. The structure of visuospatial memory in adulthood. Learning and Individual Differences 2013 Jun;25:99-110. [doi: 10.1016/j.lindif.2013.01.014]
44. Hultsch DF, MacDonald SW, Hunter MA, Maitland SB, Dixon RA. Sampling and generalisability in developmental research: Comparison of random and convenience samples of older adults. International Journal of Behavioral Development 2016 Jun 30;26(4):345-359. [doi: 10.1080/01650250143000247]

Abbreviations

ADD: Alzheimer disease dementia
AI: artificial intelligence
GDS: Geriatric Depression Scale
MAU: monthly active user
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
MMT: metamemory training
RI: Rappel Indicé
rmANOVA: repeated-measures analysis of variance
SGDS: Short-Form Geriatric Depression Scale
SLDS: spoken language dialogue system
SMC: subjective memory complaints
SMCQ: Subjective Memory Complaints Questionnaire

Edited by G Eysenbach; submitted 13.05.20; peer-reviewed by N Khalili-Mahani, MDG Pimentel; comments to author 28.06.20; revised version received 20.09.20; accepted 09.01.21; published 16.02.21.

Please cite as:
Kim J, Shin E, Han K, Park S, Youn JH, Jin G, Lee JY
Efficacy of Smart Speaker–Based Metamemory Training in Older Adults: Case-Control Cohort Study
J Med Internet Res 2021;23(2):e20177
URL: http://www.jmir.org/2021/2/e20177/
doi:10.2196/20177
PMID:33591276

Original Paper

Use of Physiological Data From a Wearable Device to Identify SARS-CoV-2 Infection and Symptoms and Predict COVID-19 Diagnosis: Observational Study

Robert P Hirten[1,2], MD; Matteo Danieletto[2,3], PhD; Lewis Tomalin[4], PhD; Katie Hyewon Choi[4], MS; Micol Zweig[2,3], MPH; Eddye Golden[2,3], MPH; Sparshdeep Kaur[2], BBA; Drew Helmus[1], MPH; Anthony Biello[1], BA; Renata Pyzik[5], MS; Alexander Charney[3,6,7], MD; Riccardo Miotto[2,3], PhD; Benjamin S Glicksberg[2,3], PhD; Matthew Levin[8], MD; Ismail Nabeel[9], MD; Judith Aberg[10], MD; David Reich[8], MD; Dennis Charney[11,12], MD; Erwin P Bottinger[2], MD; Laurie Keefer[1,6], PhD; Mayte Suarez-Farinas[3,4], PhD; Girish N Nadkarni[2,13,14], MD; Zahi A Fayad[5,15], PhD

[1]The Dr Henry D Janowitz Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[2]Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[4]Center for Biostatistics, Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[5]The BioMedical Engineering and Imaging Institute, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[6]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[7]Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[8]Department of Anesthesiology, Perioperative and Pain Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[9]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[10]Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[11]Office of the Dean, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[12]Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, United States

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

[14]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[15]Department of Diagnostic, Molecular and Interventional Radiology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

Corresponding Author:
Robert P Hirten, MD
The Dr Henry D Janowitz Division of Gastroenterology
Icahn School of Medicine at Mount Sinai
1468 Madison Avenue, Annenberg Building RM 5-12
New York, NY, 10029
United States
Phone: 1 212 241 0150
Email: robert.hirten@mountsinai.org

Abstract

Background: Changes in autonomic nervous system function, characterized by heart rate variability (HRV), have been associated with infection and observed prior to its clinical identification.

Objective: We performed an evaluation of HRV collected by a wearable device to identify and predict COVID-19 and its related symptoms.

Methods: Health care workers in the Mount Sinai Health System were prospectively followed in an ongoing observational study using the custom Warrior Watch Study app, which was downloaded to their smartphones. Participants wore an Apple Watch for the duration of the study, measuring HRV throughout the follow-up period. Surveys assessing infection and symptom-related questions were obtained daily.

Results: Using a mixed-effect cosinor model, the mean amplitude of the circadian pattern of the standard deviation of the interbeat interval of normal sinus beats (SDNN), an HRV metric, differed between subjects with and without COVID-19 (P=.006). The mean amplitude of this circadian pattern differed between individuals during the 7 days before and the 7 days after a COVID-19

diagnosis compared to this metric during uninfected time periods (*P*=.01). Significant changes in the mean and amplitude of the circadian pattern of the SDNN was observed between the first day of reporting a COVID-19–related symptom compared to all other symptom-free days (*P*=.01).

Conclusions: Longitudinally collected HRV metrics from a commonly worn commercial wearable device (Apple Watch) can predict the diagnosis of COVID-19 and identify COVID-19–related symptoms. Prior to the diagnosis of COVID-19 by nasal swab polymerase chain reaction testing, significant changes in HRV were observed, demonstrating the predictive ability of this metric to identify COVID-19 infection.

(J Med Internet Res 2021;23(2):e26107) doi:10.2196/26107

KEYWORDS

wearable device; COVID-19; identification; prediction; heart rate variability; physiological; wearable; app; data; infectious disease; symptom; prediction; diagnosis; observational

Introduction

COVID-19 has resulted in over 41 million infections and more than 1.1 million deaths [1]. The prolonged incubation period and variable symptomatology of SARS-CoV-2 have facilitated disease spread. Approximately 30%-45% of individuals infected with SARS-CoV-2 are asymptomatic and testing is generally being limited to symptomatic individuals [2-4]. Health care workers, characterized as any type of worker in a health care system, represent a vulnerable population, with a threefold increased risk of infection compared to the general population [5]. This increased risk of transmission is important in health care settings, where asymptomatic or presymptomatic health care workers can shed the virus, contributing to transmission within health care facilities and their own households [6].

Digital health technology offers an opportunity to address the limitations of traditional public health strategies aimed at curbing the spread of COVID-19 [7]. Smartphone apps are effective in using symptoms to identify people who may be infected with SARS-CoV-2; however, these apps rely on ongoing participant compliance and self-reported symptoms [8]. Wearable devices are commonly used for remote sensing, and they provide a means to objectively quantify physiological parameters, including heart rate, sleep, activity, and measures of autonomic nervous system (ANS) function (eg, heart rate variability [HRV]) [9]. The addition of physiological data from wearable devices to symptom-tracking apps has been shown to increase the ability to identify people infected with SARS-CoV-2 [10].

HRV is a physiological metric that provides insight into the interplay between the parasympathetic and sympathetic nervous systems that modulate cardiac contractility and cause variability in the beat-to-beat intervals [11]. HRV exhibits a 24-hour circadian pattern, with relative sympathetic tone during the day and parasympathetic activity at night [12-14]. Changes in this circadian pattern can be leveraged to identify different physiological states. Several studies have demonstrated that lower HRV, indicating increased sympathetic balance, is a reliable predictor of infection onset [15,16]. However, HRV and its dynamic changes over time have not been evaluated as a marker or predictor of COVID-19. In response to the COVID-19 pandemic, we launched the Warrior Watch Study, employing a novel smartphone app to remotely enroll and monitor health care workers throughout the Mount Sinai Health System in New York City, a site of initial case surge. This digital platform enables the delivery of remote surveys to Apple iPhones and passive collection of Apple Watch data, including HRV. The aim of this study is to determine if SARS-CoV-2 infections can be identified and predicted prior to a positive test result using the longitudinal changes in HRV metrics derived from individuals' Apple Watch data.

Methods

Study Design

The primary aim of the study was to determine whether changes in HRV can differentiate participants who are infected and not infected with SARS-CoV-2. The secondary aim was to observe if changes in HRV can predict the development of SARS-CoV-2 infection prior to diagnosis by a SARS-CoV-2 nasal swab polymerase chain reaction (PCR) test. The exploratory aims were to (1) determine whether changes in HRV can identify the presence of COVID-19–related symptoms; (2) determine whether changes in HRV can predict the development of COVID-19–related symptoms; and (3) evaluate how HRV changed throughout the infection and symptom period.

Health care workers in the Mount Sinai Health System were enrolled in an ongoing prospective observational cohort study. Eligible participants were aged ≥18 years, were current employees in the Mount Sinai Health System, had an iPhone Series 6 or higher, and had or were willing to wear an Apple Watch Series 4 or higher. Participants were excluded if they had an underlying autoimmune disease or were taking medications known to interfere with ANS function. A positive COVID-19 diagnosis was defined as a positive SARS-CoV-2 nasal swab PCR test reported by the participant. Daily symptoms were collected, including fever and chills, feeling tired or weak, body aches, dry cough, sneezing, runny nose, diarrhea, sore throat, headache, shortness of breath, loss of smell or taste, itchy eyes, none, or other. This study was approved by the Institutional Review Board at The Icahn School of Medicine at Mount Sinai.

Study Procedures

Participants downloaded the custom Warrior Watch app to complete eligibility questionnaires and sign an electronic consent form. Participants completed an app-based baseline assessment collecting demographic information, prior COVID-19 diagnosis history, occupation, and medical history and were then followed

prospectively through the app. Daily survey questionnaires captured COVID-19–related symptoms, symptom severity, SARS-CoV-2 nasal swab PCR test results, serum SARS-CoV-2 antibody test results, and daily patient care–related exposure (Table S1 in Multimedia Appendix 1). Participants performed their normal activities throughout the study and were instructed to wear the Apple Watch for a minimum duration of 8 hours per day.

Wearable Monitoring Device and Autonomic Nervous System Assessment

HRV was measured via the Apple Watch Series 4 or 5, which are commercially available wearable devices. Participants wore the device on the wrist and connected it via Bluetooth to their iPhone. The Apple Watch is equipped with an enhanced photoplethysmogram optical heart sensor that combines a green light-emitting diode paired with a light-sensitive photodiode generating time series peaks that correlate with the magnitude of change in the green light generated from each heartbeat [17]. Data are filtered for ectopic beats and artifacts. The time difference between heartbeats is classified as the interbeat interval (IBI), from which HRV is calculated. The Apple Watch and the Apple Health app automatically calculate HRV using the standard deviation of the IBI of normal sinus beats (SDNN), measured in milliseconds. This time domain index reflects both sympathetic and parasympathetic nervous system activity and is calculated by the Apple Watch during ultra–short-term recording periods of approximately 60 seconds [11]. The Apple Watch generates several HRV measurements throughout a 24-hour period. HRV metrics are stored in a locally encrypted database accessible through the iPhone Health app, which is retrieved through our custom Warrior Watch app. Data are transferred from the iPhone and Apple Watch upon completion of the e-consent form and any survey in the app. Wearable data are stored locally, enabling retrieval on days when the surveys are not completed by the participants.

Statistical Analysis

HRV Modeling

The HRV data collected through the Apple Watch were characterized by a circadian pattern, a sparse sampling over a 24-hour period, and nonuniform timing across days and participants. These characteristics bias easily derived features, including mean, maximum, and minimum, creating the need to derive methods that model the circadian rhythm of HRV. A cosinor model was used to model the daily circadian rhythm over a 24-hour period with the following nonlinear function:

$$Y(t) = M + A\cos(2\pi t/\tau + \varphi) + e_i(t) \quad \textbf{(1)}$$

where τ is the period (τ=24 h); M is the midline statistic of rhythm (MESOR), a rhythm-adjusted mean; A is the amplitude, a measure of half of the extent of variation within a day; and Φ is the acrophase, a measure of the time at which overall high values recur on each day (Figure S1 in Multimedia Appendix 1). This nonlinear model with three parameters has the advantage of being easily transformed into a linear model by recoding the time (t) in two new variables x and z as $x = \sin(2\pi t/\tau)$, $z = \sin(2\pi t/\tau)$. HRV can then be written as follows:

$$Y(t) = M + \beta x_t + \gamma z_t + e_i(t) \quad \textbf{(2)}$$

where the linear coefficients β, γ of the linear model in Equation 2 are related to the nonlinear parameters of the nonlinear model in Equation 1 by $\beta = A\sin(\varphi)$ and $\gamma = -A\sin(\varphi)$. One can estimate the linear parameters β, γ and then obtain A and φ as

We took advantage of the longitudinal structure of the data to identify a participant-specific daily pattern, and we then measured departures from this pattern as a function of COVID-19 diagnosis or other relevant covariates. To do this, we used a mixed-effect cosinor model, where the HRV measure of participant *i* at time *t* can be written as follows:

$$HRV_{it} = (M + \beta.x_{it} + \gamma.z_{it}) + W_{it}.\theta_i + e_i(t),\ e_i(t) \sim N(0,s) \quad \textbf{(5)}$$

where M, β, and γ are the population parameters (fixed effects) and θ_i is a vector of random effects that is assumed to follow a multivariate normal distribution $\theta_i \sim N(0,\Sigma)$. In this context, the introduction of random effects intrinsically models the correlation due to the longitudinal sampling. To measure the impact of any covariate C on the participants' daily curve, we can introduce these covariates as fixed effects, as their interactions with x and z:

$$HRV_{it} = M + a_O C_i + (\beta + a_2 C_i).x_{it} + (\gamma + a_3 C_i).z_{it} + W_{it}.\theta_i + e_i(t) \quad \textbf{(6)}$$

The model parameters and the standard errors of Equation 6 can be estimated via maximum likelihood or reweighted least squares (REWL), and hypothesis testing can be conducted for any comparison that can be written as a linear function of the *a*, β, and γ parameters.

However, to test if the cosinor curve, defined by the nonlinear parameters M, A, and Φ in Equation 1, differs between the populations defined by the covariate C, we proposed a bootstrapping procedure in which for each resampling iteration, we (1) fit a linear mixed-effect model using REWL; (2) estimated the marginal means by obtaining the linear parameters for each group defined by covariate C; (3) used the inverse relationship to estimate marginal means M, A, and Φ for each group defined by C; and (4) defined the bootstrapping statistics as the pairwise differences of M, A, and Φ between groups defined by C. For these iterations, the confidence intervals for the nonlinear parameter were defined using standard bootstrap techniques, and the *P* values derived for the differences in each nonlinear parameter between groups were defined by C_i. Age and sex were included as covariates in the HRV analyses and admitted as invariant and time-variant covariates.

Association and Prediction of COVID-19 Diagnosis and Symptoms

The relationship between a COVID-19 diagnosis and the change in the HRV curves was evaluated. To test this association, we defined the time variant covariate C_{it} for participant i at time t as follows:

HRV metrics for the 14 days following the time of the first positive SARS-CoV-2 nasal swab PCR test were used to define the positive SARS-CoV-2 infection window. To evaluate the predictive ability of changes in HRV prior to a COVID-19 diagnosis and to explore its changes during the infection period, the time variant covariate was used to characterize the following 4 groups: healthy uninfected individuals ($t<t_0-7$), 7 days before COVID-19 diagnosis ($t\geq t_0-7$, $t<t_0$), the first 7 days post–COVID-19 diagnosis ($t_0\leq t<t_0+7$), and 7-14 days postdiagnosis ($t_0+7\leq t<t_0+14$).

To determine the association between COVID-19 symptoms and changes in HRV metrics, we defined being symptomatic as the first day of a reported symptom and compared this to all other days. To evaluate the predictive ability of HRV to identify upcoming symptom days and to explore its changes over time, the time variant covariate was used to characterize the following 4 groups: healthy asymptomatic individuals for $t<t_0-1$, 1 day before COVID-19 symptoms ($t\geq t_0-1$, $t<t_0$), the first day of COVID-19 symptoms ($t_0\leq t<t_0+1$) and 1 day post–COVID-19 symptom development ($t_0+1\leq t<t_0+2$).

Results

Participant Demographics

We enrolled 297 participants between April 29 and September 29, 2020, when the data were censored for analysis (Table 1). The median age at enrollment was 36 years (SD 9.8), and 69.4% of participants (204/297) were women. Of the 297 participants, 20 (6.7%) reported having a positive SARS-CoV-2 nasal swab PCR test prior to enrollment, while 28 participants (9.4%) reported having a positive blood antibody test prior to joining the study. The median duration of follow-up was 42 days (range 0-152 days). A median of 28 HRV samples (range 1-129) were obtained per participant. Study compliance over the follow-up period, defined as participants answering over 50% of daily surveys, was 70.4%.

Table 1. Baseline demographics of the study participants at enrollment (N=297).

Characteristic	Value
Age (years), mean (SD)	36.3 (9.8)
BMI (kg/m^2), mean (SD)	25.6 (5.7)
Female gender, n (%)	204 (69.4)
Race, n (%)	
Asian	73 (24.6)
Black	29 (9.8)
White	108 (36.4)
Other	43 (14.5)
Hispanic ethnicity, n (%)	44 (14.8)
Baseline positive SARS-CoV-2 nasal swab PCR[a] test, n (%)	20 (6.7)
Baseline positive SARS-CoV-2 serum antibody test, n (%)	28 (9.4)
Occupation[b], n (%)	
Clinical nontrainee	198 (68.0)
Clinical trainee	36 (12.4)
Nonclinical staff	57 (19.6)
Baseline smoking status, n (%)	
Current or past smoker	35 (11.9)
Nonsmoker or rare smoker	259 (88.1)
Baseline immune suppressing medication, n (%)	4 (1.4)

[a]PCR: polymerase chain reaction.

[b]Clinical trainee is defined as a resident or fellow; clinical nontrainee is defined as a health care worker reporting at least one patient-facing day during follow-up, exclusive of residents and fellows; nonclinical staff is defined as a health care worker who did not report a patient-facing day during follow-up.

Identification and Prediction of COVID-19 Diagnosis

Participants classified as not having a COVID-19 diagnosis during follow-up included those with and without a diagnosis of COVID-19 prior to study enrollment. There was no significant difference in the mean MESOR, acrophase, or amplitude of the circadian SDNN pattern of participants with a positive nasal swab PCR test prior to enrollment compared to those who were never diagnosed with COVID-19. This supports the inclusion of participants with a prior COVID-19 diagnosis

in our analysis (Table S2 in Multimedia Appendix 1). During the follow-up period, 13/297 participants (4.4%) reported a positive SARS-CoV-2 nasal swab PCR test, with the date of diagnosis corresponding with the reported date of the positive test. The mean MESOR, acrophase, and amplitude of the circadian SDNN pattern in participants who were and were not diagnosed with COVID-19 during follow-up are described in Table 2. A significant difference in the circadian pattern of SDNN was observed in participants diagnosed with COVID-19 compared to those without a COVID-19 diagnosis. There was a significant difference (*P*=.006) between the mean amplitude of the circadian pattern of SDNN in participants with COVID-19 (1.23 milliseconds, 95% CI –1.94 to 3.11) and without COVID-19 (5.30 milliseconds, 95% CI 4.97 to 5.65). No difference was observed between the MESOR (*P*=.46) or acrophase (*P*=.80) in these two infection states (Figure 1A-C). Similar findings were observed when this analysis was repeated to include only participants who had either a positive (13/297, 4.4%) or negative (108/297, 36.4%) SARS-CoV-2 nasal swab PCR test during the follow-up period, excluding participants who reported never being tested (Table S3 in Multimedia Appendix 1).

The mean MESOR, acrophase, and amplitude of the circadian SDNN pattern for those without COVID-19, those during the 7 days prior to a COVID-19 diagnosis, participants during the 7 days after a COVID-19 diagnosis, and those during the 7-14 days after a COVID-19 diagnosis are described in Table 3. Significant changes in the circadian SDNN pattern were observed in participants during the 7 days prior to and the 7 days after a diagnosis of COVID-19 when compared to uninfected participants. There was a significant difference between the amplitude of the SDNN circadian rhythm between uninfected participants (5.31 milliseconds, 95% CI 4.95 to 5.67) compared to individuals during the 7-day period prior to a COVID-19 diagnosis (0.29 milliseconds, 95% CI –4.68 to 1.73; *P*=.01) and participants during the 7 days after a COVID-19 diagnosis (1.22 milliseconds, 95% CI –2.60 to 3.25; *P*=.01). There were no other significant differences between the MESOR, amplitude, and acrophase of the circadian rhythm of the SDNN observed between healthy individuals, individuals 7 days before a COVID-19 diagnosis, individuals 7 days after a COVID-19 diagnosis, and individuals 7-14 days after infection (Figure 1D-E).

Of the 13 subjects diagnosed with COVID-19 during follow up, 6 reported symptoms at some point during the study period. Only 4 subjects had symptomatic COVID-19 infections, reporting symptoms between 7 days prior and 14 days after a positive SARS-CoV-2 nasal swab PCR test. Comparing participants with and without symptomatic COVID-19, no significant differences between the MESOR (28.58 milliseconds, 95% CI 18.61 to 38.56; 37.71 milliseconds, 95% CI 30.65 to 44.98, *P*=.11), amplitude (1.15 milliseconds, 95% CI –2.63 to 3.21; 1.68 milliseconds, 95% CI –1.13 to 3.95, *P*=.76) and acrophase (–1.92 milliseconds, 95% CI –3.68 to –0.02; –2.49 milliseconds, 95% CI –4.37 to –0.33, *P*=.62) of the circadian rhythm of SDNN were observed, respectively.

Table 2. HRV parameters in participants with and without COVID-19 diagnoses based on SARS-CoV-2 nasal swab PCR tests.

Parameter	Parameter (milliseconds), mean (95% CI)		Difference (95% CI)	*P* value
	Participants not diagnosed with COVID-19	Participants diagnosed with COVID-19		
MESOR[a]	43.57 (41.40 to 45.40)	42.46 (38.90 to 45.79)	–1.12 (–4.22 to 1.73)	.46
Amplitude	5.30 (4.97 to 5.65)	1.23 (–1.94 to 3.11)	–4.07 (–7.29 to –2.07)	*.006* [b]
Acrophase	–2.44 (–2.49 to –2.39)	–2.23 (–2.22 to –4.24)	0.22 (–1.74 to 2.43)	.80

[a]MESOR: midline statistic of rhythm.

[b]Italic text indicates statistical significance.

Figure 1. Relationship between HRV circadian rhythm and COVID-19 status. Timeline (A) illustrates HRV measures from the time of COVID-19 diagnosis via nasal swab PCR test and during the following 2 weeks after subjects were deemed to be COVID-19–positive and were compared with measurements outside this window, where subjects were deemed to be COVID-19–negative. Daily HRV rhythm (B) on days with positive and negative COVID-19 diagnoses. Plots (C) showing the means and 95% confidence intervals for the parameters defining the circadian rhythm, acrophase, amplitude and MESOR, on days with positive and negative COVID-19 diagnoses. Daily HRV patterns (D, E) for days on which subjects were healthy, 7 days before a positive COVID-19 test, 7 days after a positive COVID-19 test, and 7-14 days after a positive COVID-19 test. Means and 95% confidence intervals for the acrophase, amplitude, and MESOR of the HRV measured on days when participants were healthy, 7 days before a positive COVID-19 test, 7 days after a positive COVID-19 test, and 7-14 days after a positive COVID-19 test. *P<.10, **P<.05, ***P<.01, ****P<.001, ns: not significant. HRV: heart rate variability; MESOR: midline statistic of rhythm; SDNN: standard deviation of the interbeat interval of normal sinus beats.

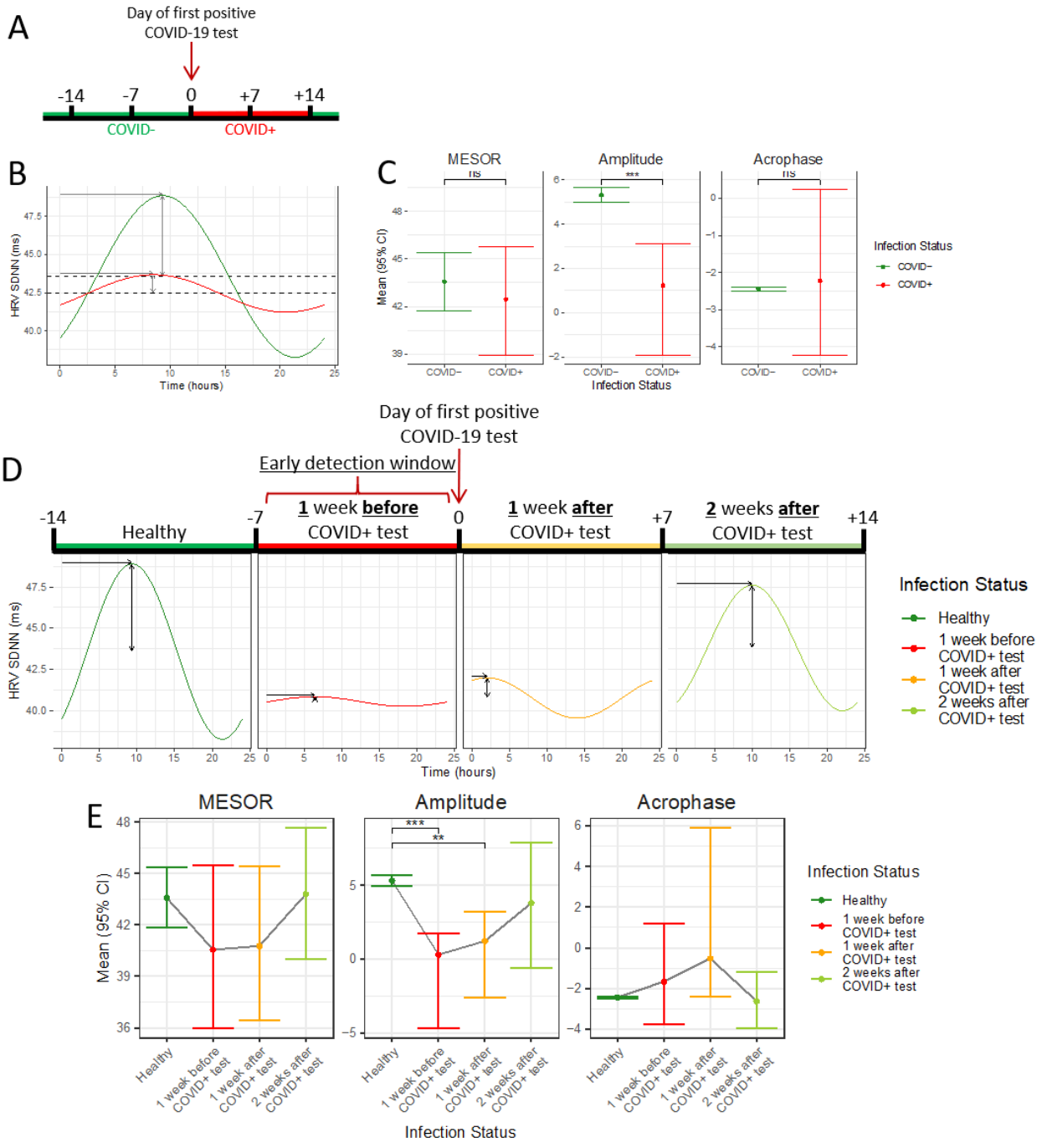


Table 3. Comparison of heart rate variability parameters based on the time periods before and after diagnosis.

Parameter and first period relative to COVID-19 diagnosis	Value (milliseconds), mean (95% CI)	Second period relative to COVID-19 diagnosis	Value (milliseconds), mean (95% CI)	Difference (95% CI)	*P* value
MESOR[a]					
7 days before	40.56 (35.98 to 45.46)	Uninfected	43.58 (41.88 to 45.37)	–3.03 (–6.98 to 1.02)	.13
7 days after	40.77 (36.44 to 45.42)	Uninfected	43.58 (41.88 to 45.37)	–2.81 (–6.73 to 1.10)	.17
7-14 days after	43.80 (40.01 to 47.65)	Uninfected	43.58 (41.88 to 45.37)	0.22 (–3.39 to 3.73)	.89
7 days before	40.56 (35.98 to 45.46)	7-14 days after	43.80 (40.01 to 47.65)	–3.24 (–9.63 to 3.33)	.32
7 days after	40.77 (36.44 to 45.42)	7-14 days after	43.80 (40.01 to 47.65)	–3.03 (–6.98 to 1.02)	.13
7 days after	40.77 (36.44 to 45.42)	7 days before	40.56 (35.98 to 45.46)	0.217 to (–3.39 to 3.73)	.89
Amplitude					
7 days before	0.29 (–4.68 to 1.73)	Uninfected	5.31 (4.95 to 5.67)	–5.02 (–10.14 to –3.58)	*.01* [b]
7 days after	1.22 (–2.60 to 3.25)	Uninfected	5.31 (4.95 to 5.67)	–4.09 (–7.87 to –1.93)	*.01*
7-14 days after	3.80 (–0.64 to 7.88)	Uninfected	5.31 (4.95 to 5.67)	–1.51 (–5.79 to 2.35)	.48
7 days before	0.29 (–4.68 to 1.73)	7-14 days after	3.80 (–0.64 to 7.88)	–3.51 (–10.50 to 0.22)	.20
7 days after	1.22 (–2.60 to 3.25)	7-14 days after	3.80 (–0.64 to 7.88)	–2.58 (–8.44 to 2.08)	.34
7 days after	1.22 (–2.60 to 3.25)	7 days before	0.29 (–4.68 to 1.73)	0.93 (–1.92 to 5.83)	.58
Acrophase					
7 days before	–1.67 (–3.78 to 1.19)	Uninfected	–2.44 (–2.49 to –2.39)	0.78 (–1.4 to 3.62)	.45
7 days after	–0.53 (-2.39-5.89)	Uninfected	–2.44 (–2.49 to –2.39)	1.92 (0.03 to 8.13)	.48
7-14 days after	–2.63 (–3.95 to 1.19)	Uninfected	–2.44 (–2.49 to –2.39)	–0.19 (–1.39 to 1.16)	.70
7 days before	–1.67 (–3.78 to 1.19)	7-14 days after	–2.63 (–3.95 to 1.19)	0.96 (–1.85 to 4.32)	.55
7 days after	–0.53 (–2.39 to 5.89)	7-14 days after	–2.63 (–3.95 to 1.19)	2.10 (0.10 to 8.29)	.35
7 days after	–0.53 (–2.39 to 5.89)	7 days before	–1.67 (–3.78 to 1.19)	1.14 (–1.34 to 7.27)	.58

[a]MESOR: midline statistic of rhythm.

[b]Italic text indicates statistical significance.

Identification and Prediction of COVID-19 Symptoms

Of the 297 participants, 165 (55.6%) reported developing a symptom during the follow-up period, with the greatest number of participants reporting feeling tired or weak (n=87, 29.3%), followed by headaches (n=82, 27.6%) and sore throat (n=60, 20.2%) (Table 4). Evaluating the days on which participants experienced symptoms, we found that loss of smell or taste was reported the most frequently, with a mean of 138 days. This was followed by feeling tired or weak, reported for a mean of 25 days, and runny nose, reported for a mean of 19.5 days (Figure 2).

The mean MESOR, acrophase, and amplitude observed in the circadian SDNN patterns of participants on the first day they experienced a symptom and on all other days of follow-up are reported in Table 5. There was a significant difference in the circadian SDNN pattern between participants on the first day a symptom was reported compared to all other days of follow-up. Specifically, there was a significant difference (*P*=.01) between the mean MESOR of the circadian SDNN pattern on the first day of symptoms (46.01 milliseconds, 95% CI 43.37 to 48.77) compared to all other days (43.48 milliseconds, 95% CI 41.77 to 45.27). Similarly, there was a significant difference (*P*=.01) between the mean amplitude of the circadian SDNN pattern on the first day of symptoms (2.58 milliseconds, 95% CI 0.26-5.00) compared to all other days (5.30 milliseconds, 95% CI 4.95-5.66) (Figure 3A-C). Out of the 165 participants reporting symptoms during the follow-up period, 36 participants (21.2%) reported experiencing a symptom graded as 6 or higher on a 10-point scale. The impact of the severity of a symptom on the HRV was evaluated by comparing HRV metrics in participants with symptoms graded as a 6 or higher versus those with symptoms graded 5 or less. There was a significant difference between the mean amplitude (*P*=.02) and MESOR (*P*=.01) of the circadian SDNN pattern in subjects with low symptom severity (amplitude 9.39 milliseconds, 95% CI 7.41 to 11.02; MESOR 42.82 milliseconds, 95% CI 38.65 to 47.19) and high symptom severity (amplitude 4.74 milliseconds, 95% CI 0.75 to 8.40; MESOR 36.15 milliseconds, 95% CI 29.11 to 43.34). There was no significant difference (*P*=.84) in the acrophase between those with low symptom severity (–2.43 ms, 95% CI –2.59 to –2.26) and high symptom severity (–2.49 milliseconds, 95% CI –3.19 to –1.77).

The mean MESOR, acrophase, and amplitude observed in the circadian SDNN patterns of participants on the day before symptoms developed, on the first day of the symptom, on the

day following the first day of the symptom, and on all other days are reported in Table 6. Significant changes in the circadian SDNN pattern were observed, specifically in the mean amplitude (*P*=.04), when comparing participants on the first day of the symptom (3.07 milliseconds, 95% CI 0.88 to 5.22) to all other days (5.32 milliseconds, 95% CI 4.99 to 5.66). Excluded from this analysis were the day prior to and the day after the first symptomatic day. Changes in SDNN characteristics trended toward significance prior to the development of symptoms. Specifically, the differences in the mean amplitude of the circadian SDNN pattern trended toward significance when comparing the day prior to symptom development (2.92 milliseconds, 95% CI 0.50 to 5.33) with all other days (5.32 milliseconds, 95% CI 4.99 to 5.66; *P*=.06). Again, the first day of the symptom and the day after the first symptomatic day were excluded from the analysis. Additionally, there was a trend toward significance when comparing the amplitude of the SDNN circadian pattern between participants during the first day of the symptom (3.07 milliseconds, 95% CI 0.88 to 5.22) with that one day after the first symptom was reported (5.47 milliseconds, 95% CI 3.16 to 7.76; *P*=.56). Excluded from the analysis were the day prior to symptom development and all other days. There were no other significant differences between the MESOR, amplitude, or acrophase of the circadian rhythm of the SDNNs when comparing participants on the day before symptoms developed, on the first day of the symptom, on the day following the first day of the symptom, or on all other days (Figure 3D-E).

Table 4. Number of participants reporting each symptom (N=297).

Symptom	Participants, n (%)[a]
Fever or chills	11 (3.7)
Fatigue or weakness	87 (29.3)
Body aches	47 (15.8)
Dry cough	32 (10.8)
Sneezing	52 (17.5)
Runny nose	43 (14.4)
Diarrhea	33 (11.1)
Sore throat	60 (20.2)
Headache	82 (27.6)
Shortness of breath	11 (3.7)
Loss of smell or taste	5 (1.7)
Itchy eyes	53 (17.8)
Other	26 (8.8)

[a]Percentages add to >100% because participants could report one or more symptoms.

Figure 2. Number of symptom days per participant when evaluating days on which participants reported symptoms.

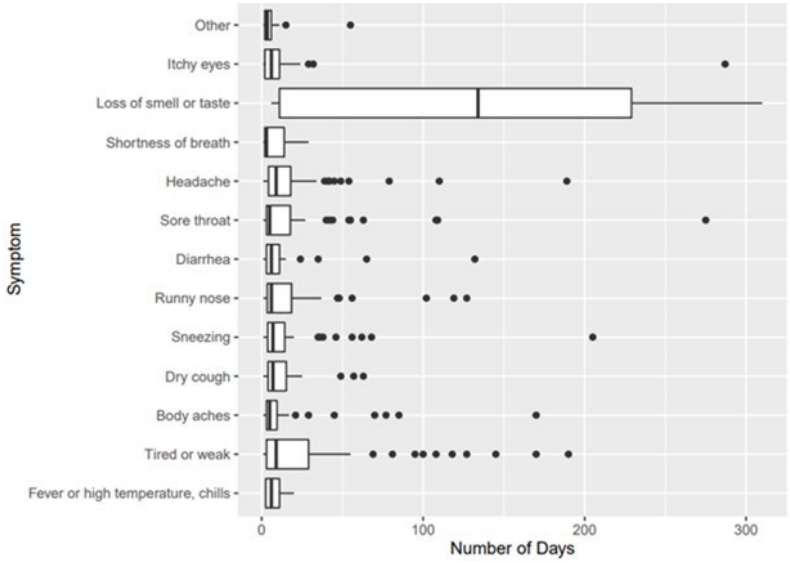


Table 5. Heart rate variability parameters on the first day of reported symptoms compared to all other symptom-free days.

Parameter	Value on the first day of symptoms (milliseconds), mean (95% CI)	Value on all other days (milliseconds), mean (95% CI)	Difference (95% CI)	*P* value
MESOR[a]	46.01 (43.37 to 48.77)	43.48 (41.77 to 45.27)	2.53 (0.82 to 4.36)	*.01* [b]
Amplitude	2.58 (0.26 to 5.00)	5.30 (4.95 to 5.66)	–2.73 (–5.16 to 0.31)	*.01*
Acrophase	–2.21 (–2.83 to –1.58)	–2.44 (–2.49 to –2.39)	0.24 (–0.38 to 0.88)	.44

[a]MESOR: midline statistic of rhythm.

[b]Italic text indicates statistical significance.

Figure 3. Relationship between HRV circadian rhythm and symptom onset. Timeline (A) illustrates the timing of symptom onset; the HRV profiles of the day of the first symptom are compared to all other days. Daily HRV rhythm (B) on the day of the first symptom and nonsymptomatic or late-symptom days. Plots (C) showing means and 95% confidence intervals for the parameters defining the circadian rhythm, acrophase, amplitude, and MESOR, on first symptom and nonsymptomatic or late-symptomatic days. Daily HRV pattern (D) for nonsymptomatic or late-symptomatic days, the day before the first symptom, the day of the first symptom, and the day after the first symptom. Means and 95% confidence intervals for the acrophase, amplitude, and MESOR of the HRV measured on nonsymptomatic or late-symptomatic days, the day before the first symptom, the day of the first symptom, and the day after the first symptom. **P*<.10, ***P*<.05, ****P*<.01, *****P*<.001, ns, not significant. HRV: heart rate variability; MESOR: midline statistic of rhythm; SDNN: standard deviation of the interbeat interval of normal sinus beats.

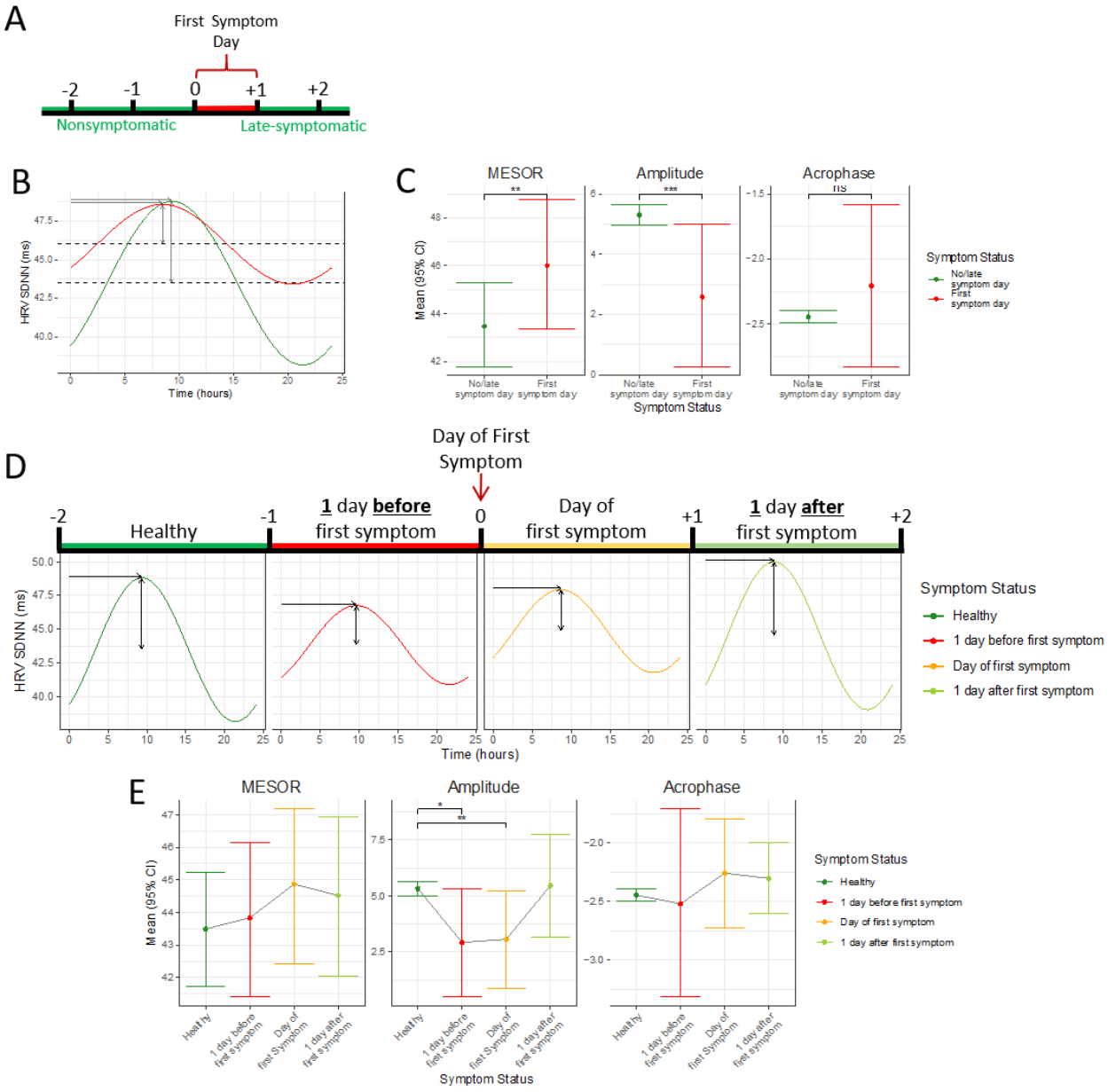


Table 6. Comparison of heart rate variability parameters based on symptom state and the time periods before and after the first day of reported symptoms.

Parameter and first symptom state	Value (milliseconds), mean (95% CI)	Second symptom state	Value (milliseconds), mean (95% CI)	Difference (95% CI)	*P* value
MESOR[a]					
1 day after first symptom day	44.52 (42.05 to 46.94)	Asymptomatic	43.49 (41.74 to 45.21)	1.03 (–0.64 to 2.67)	.21
1 day before first symptom day	43.84 (41.41 to 46.15)	Asymptomatic	43.49 (41.74 to 45.21)	0.34 (–1.46 to 2.23)	.73
First day of symptom	44.87 (42.42 to 47.18)	Asymptomatic	45.49 (41.74 to 45.21)	1.37 (–0.24 to 3.04)	.11
1 day before first symptom day	43.84 (41.41 to 46.15)	1 day after first symptom day	44.52 (42.05 to 46.94)	–0.69 (–3.72 to 2.47)	.66
First day of symptom	44.87 (42.42 to 47.18)	1 day after first symptom day	44.52 (42.05 to 46.94)	0.34 (–1.46 to 2.23)	.73
First day of symptom	44.87 (42.42 to 47.18)	1 day before first symptom day	43.84 (41.41 to 46.15)	1.03 (–0.64 to 2.67)	.21
Amplitude					
1 day after first symptom day	5.47 (3.16 to 7.76)	Asymptomatic	5.32 (4.99 to 5.66)	0.15 (–2.21 to 2.37)	.91
1 day before first symptom day	2.92 (0.50 to 5.33)	Asymptomatic	5.32 (4.99 to 5.66)	–2.40 (–4.75 to –0.07)	.06
First day of symptom	3.07 (0.88 to 5.22)	Asymptomatic	5.32 (4.99 to 5.66)	–2.25 (–4.38 to –0.27)	*.04* [b]
1 day before first symptom day	2.92 (0.50 to 5.33)	1 day after first symptom day	5.47 (3.16 to 7.76)	–2.55 (–6.64 to 1.65)	.25
First day of symptom	3.07 (0.88 to 5.22)	1 day after first symptom day	5.47 (3.16 to 7.76)	–2.40 (–4.75 to –0.06)	.06
First day of symptom	3.07 (0.88 to 5.22)	1 day before first symptom day	2.92 (0.50 to 5.33)	0.15 (–2.20 to 2.37)	.91
Acrophase					
1 day after first symptom day	–2.30 (–2.60 to –2.00)	Asymptomatic	–2.45 (–2.50 to –2.39)	0.14 (–0.15 to 0.44)	.33
1 day before first symptom day	–2.52 (–3.31 to –1.71)	Asymptomatic	–2.45 (–2.50 to –2.39)	–0.08 (–0.79 to 0.66)	.86
First day of symptom	–2.26 (–2.73 to –1.79)	Asymptomatic	–2.45 (–2.50 to –2.39)	0.19 (–0.24 to 0.63)	.36
1 day before first symptom day	–2.52 (–3.31 to –1.71)	1 day after first symptom day	–2.30 (–2.60 to –2.00)	–0.22 (–1.11 to 0.70)	.63
First day of symptom	–2.26 (–2.73 to –1.79)	1 day after first symptom day	–2.30 (–2.60 to –2.00)	0.04 (–0.36 to 0.46)	.86
First day of symptom	–2.26 (–2.73 to –1.79)	1 day before first symptom day	–2.52 (–3.31 to –1.71)	0.26 (–0.40 to 0.92)	.41

[a]MESOR: midline statistic of rhythm.

[b]Italic text indicates statistical significance.

Discussion

Principal Results and Comparison with Prior Work

In this prospective study, longitudinally evaluated HRV metrics were found to be associated with a positive SARS-CoV-2 diagnosis and COVID-19 symptoms. Significant changes in these metrics were observed 7 days prior to the diagnosis of COVID-19. To the best of our knowledge, this is the first study to demonstrate that physiological metrics derived from a commonly worn wearable device (Apple Watch) can identify and predict SARS-CoV-2 infection prior to diagnosis with a SARS-CoV-2 nasal swab PCR test. These preliminary results identify a novel, easily measured physiological metric that may aid in the tracking and identification of SARS-CoV-2 infections.

Current means to control COVID-19 spread rely on case isolation and contact tracing, which have played major roles in the successful containment of prior infectious disease outbreaks [18-20]. However, due to the variable incubation period, high percentage of asymptomatic carriers, and infectivity during the presymptomatic period of COVID-19, containment of the disease has been challenging [21]. This has further limited the utility of systematic screening technologies reliant on vital sign assessment or self-reporting of symptoms [7]. Advances in digital health provide a unique opportunity to enhance disease containment. Wearable devices are commonly used and well accepted for health monitoring [9,22]. Commercially available devices are able to continually collect several physiological parameters. Unlike app-based platforms, wearable devices have the advantage of not requiring users to actively participate aside from regular use of the device. Prior to the COVID-19

pandemic, population-level data from the Fitbit wearable device demonstrated effectiveness of real-time geographic surveillance of influenza-like illnesses through the assessment of physiological parameters [23]. This concept was recently expanded during the COVID-19 pandemic by Quer and colleagues [10], who demonstrated that the combination of symptom-based data with resting heart rate and sleep data from wearable devices was superior to relying on symptom-based data alone to identify COVID-19 infections.

HRV has been shown to be altered during illnesses, with several small studies demonstrating changes in HRV associated with and predictive of the development of infection [24]. Ahmad and colleagues [16] followed 21 subjects undergoing bone marrow transplant, and they found a significant reduction in root mean square successive difference metrics prior to the clinical diagnosis of infection. Furthermore, wavelet HRV was noted to decrease by 25% on average 35 hours prior to a diagnosis of sepsis in 14 patients. In another study in 100 infants [15], significant HRV changes were noted 3-4 days preceding sepsis or systemic inflammatory response syndrome, with the largest increase being seen 24 hours prior to development. Building on these observations demonstrating that ANS changes accompany or precede infection, our team launched the Warrior Watch Study.

We demonstrated that significant changes in the circadian pattern of HRV, specifically the amplitude of SDNN, were associated with a positive COVID-19 diagnosis. Interestingly, when we compared these changes over the 7 days preceding the diagnosis of COVID-19, we continued to see significant alterations in amplitude when compared to individuals without COVID-19. This demonstrates the predictive ability of this metric to identify infection. Additionally, most participants diagnosed with COVID-19 in our cohort were asymptomatic. We demonstrated that there was no difference in changes in HRV metrics between participants with and without symptomatic COVID-19 infections. These findings support the utility of using wearable technology to identify COVID-19 infections, even in asymptomatic individuals. When we followed individuals 7-14 days after diagnosis with COVID-19, we found that the circadian HRV pattern began to normalize and was no longer statistically different from an uninfected pattern. As an exploratory analysis, we evaluated how HRV was impacted by symptoms associated with a COVID-19 diagnosis, as individuals may not be tested despite experiencing symptoms. We found significant changes in the amplitude of the circadian HRV pattern on the first day of symptoms, with a trend toward statistical significance on the days before and after symptoms were reported. Similarly, we found significant changes in HRV when we stratified subjects based on severe or nonsevere symptom severity. Taken together, these findings highlight the possible use of HRV collected via wearable devices to identify and predict COVID-19 infection.

Limitations

There are several limitations to our study. First, the number of participants who were diagnosed with COVID-19 in our cohort was small, limiting our ability to determine how predictive HRV can be of infection. However, these preliminary findings support the further evaluation of HRV as a metric to identify and predict COVID-19 and warrant further study. An additional limitation is the sporadic collection of HRV by the Apple Watch. Although our statistical modeling was able to account for this, a denser dataset would enable expanded evaluation of the relationship between this metric and infections/symptoms. The Apple Watch also only provides HRV in one time domain (SDNN), limiting assessment of the relationship between other HRV parameters with COVID-19 outcomes. Additionally, we did not capture the times of day during which participants were awake or sleeping. Therefore, fluctuations in sleep patterns may have impacted some HRV readings and could not be controlled in the analysis. Finally, an additional limitation is that we relied on self-reported data in this study, precluding independent verification of COVID-19 diagnosis.

Conclusions

In summary, we demonstrated a relationship between longitudinally collected HRV acquired from a commonly used wearable device and SARS-CoV-2 infection. These preliminary results support the further evaluation of HRV as a biomarker of SARS-CoV-2 infection by remote sensing. Although further study is needed, our findings may enable the identification of SARS-CoV-2 infection during the presymptomatic period, in asymptomatic carriers, and prior to diagnosis by a SARS-CoV-2 nasal swab PCR test. These findings warrant further evaluation of this approach to track and identify COVID-19 infections and possibly other types of infection.

Acknowledgments

Support for this study was provided by the Ehrenkranz Lab For Human Resilience, the BioMedical Engineering and Imaging Institute, The Hasso Plattner Institute for Digital Health at Mount Sinai, The Mount Sinai Clinical Intelligence Center and The Dr Henry D Janowitz Division of Gastroenterology.

Authors' Contributions

RPH, MD, GN, and ZAF developed the study concept. RPH assisted with the drafting of the manuscript. RPH, MD, LT, HC, MZ, EG, SK, DH, AB, RP, AC, RM, BG, ML, IN, DR, DC, EPB, LK, MSF, GNN, ZAF, and JA critically revised the manuscript for important intellectual content. RPH, MD, LT, HC, MZ, EG, SK, DH, AB, RP, AC, RM, BG, ML, IN, DR, DC, EPB, LK, MSF, GNN, ZAF, and JA provided final approval of the version of the manuscript to be published and agree to be accountable for all aspects of the work. All authors approved the authorship list. All authors had full access to all the data in the manuscript

and had final responsibility for the decision to submit for publication. RPH, ZAF, MSF, MD, and LT verified the underlying data.

Conflicts of Interest

RPH discloses consulting fees from HealthMode, Inc, Janssen Pharmaceuticals, and Takeda Pharmaceuticals and research support from Intralytix Inc and a Crohn's and Colitis Foundation Career Development Award (grant number 607934). BSG has received consulting fees from Data2Discovery, Sema4, and University of California San Francisco. DC is a coinventor on patents filed by the Icahn School of Medicine at Mount Sinai (ISMMS) relating to the treatment for treatment-resistant depression, suicidal ideation, and other disorders. ISMMS has entered into a licensing agreement with Janssen Pharmaceuticals, Inc, and it has received and will receive payments from Janssen under the license agreement related to these patents for the treatment of treatment-resistant depression and suicidal ideation. Consistent with the ISMMS Faculty Handbook (the medical school policy), AC is entitled to a portion of the payments received by the ISMMS. Because SPRAVATO has received regulatory approval for treatment-resistant depression, through the ISMMS, AC will be entitled to additional payments beyond those already received under the license agreement. AC is a named coinventor on several patents filed by ISMMS for a cognitive training intervention to treat depression and related psychiatric disorders. The ISMMS has entered into a licensing agreement with Click Therapeutics, Inc, and has received and will receive payments related to the use of this cognitive training intervention for the treatment of psychiatric disorders. In accordance with the ISMMS Faculty Handbook, AC has received a portion of these payments and is entitled to a portion of any additional payments that the medical school may receive from this license with Click Therapeutics. AC is a named coinventor on a patent application filed by the ISMMS for the use of intranasally administered Neuropeptide Y for the treatment of mood and anxiety disorders. This intellectual property has not been licensed. AC is a named coinventor on a patent application in the United States and several issued patents outside the United States filed by the ISMMS related to the use of ketamine for the treatment of posttraumatic stress disorder. This intellectual property has not been licensed. EPB reports consultancy agreements with Deloitte and Roland Berger; ownership interest in Digital Medicine E. Böttinger GmbH, EBCW GmbH, and Ontomics, Inc; receiving honoraria from Bayer, Bosch Health Campus, Sanofi, and Siemens; and serving as a scientific advisor or member of Bosch Health Campus and Seer Biosciences Inc. LK declares research funding from Abbvie and Pfizer, consulting for Abbvie and Pfizer, and equity ownership/stock options in MetaMe Health and Trellus Health. MSF declares research support from Novartis and Allergenis. GNN reports employment with, consultancy agreements with, and ownership interest in Pensieve Health and Renalytix AI; receiving consulting fees from AstraZeneca, BioVie, GLG Consulting, and Reata; and serving as a scientific advisor or member of Pensieve Health and Renalytix AI. ZAF discloses consulting fees from Alexion, GlaxoSmithKline, and Trained Therapeutix Discovery and research funding from Daiichi Sankyo, Amgen, Bristol Myers Squibb, and Siemens Healthineers. ZAF receives financial compensation as a board member and advisor to Trained Therapeutix Discovery and owns equity in Trained Therapeutix Discovery as a cofounder.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 117 KB - jmir_v23i2e26107_app1.docx]

References

1. Coronavirus Resource Center. Johns Hopkins University & Medicine. URL: https://coronavirus.jhu.edu/ [accessed 2020-10-22]
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
3. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 2020 May 15;26(5):672-675. [doi: 10.1038/s41591-020-0869-5] [Medline: 32296168]
4. Oran DP, Topol EJ. Prevalence of asymptomatic SARS-CoV-2 infection. Ann Intern Med 2020 Sep 01;173(5):362-367. [doi: 10.7326/m20-3012]
5. Nguyen LH, Drew DA, Graham MS, Joshi AD, Guo C, Ma W, et al. Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. Lancet Public Health 2020 Sep;5(9):e475-e483. [doi: 10.1016/S2468-2667(20)30164-X]
6. Shah ASV, Wood R, Gribben C, Caldwell D, Bishop J, Weir A, et al. Risk of hospital admission with coronavirus disease 2019 in healthcare workers and their households: nationwide linkage cohort study. BMJ 2020 Oct 28;371:m3582 [FREE Full text] [doi: 10.1136/bmj.m3582] [Medline: 33115726]
7. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health 2020 Aug;2(8):e435-e440 [FREE Full text] [doi: 10.1016/S2589-7500(20)30142-4]
8. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul 11;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]

9. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. 2020 Jan 09. URL: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2021-02-08]
10. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med 2021 Jan 29;27(1):73-77. [doi: 10.1038/s41591-020-1123-x] [Medline: 33122860]
11. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017 Sep;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
12. Boudreau P, Yeh WH, Dumont GA, Boivin DB. A circadian rhythm in heart rate variability contributes to the increased cardiac sympathovagal response to awakening in the morning. Chronobiol Int 2012 Jul 27;29(6):757-768. [doi: 10.3109/07420528.2012.674592] [Medline: 22734576]
13. Buijs RM, la Fleur SE, Wortel J, Van Heyningen C, Zuiddam L, Mettenleiter TC, et al. The suprachiasmatic nucleus balances sympathetic and parasympathetic output to peripheral organs through separate preautonomic neurons. J Comp Neurol 2003 Sep 08;464(1):36-48. [doi: 10.1002/cne.10765] [Medline: 12866127]
14. Nonell A, Bodenseh S, Lederbogen F, Kopf D, Hamann B, Gilles M, et al. Chronic but not acute hydrocortisone treatment shifts the response to an orthostatic challenge towards parasympathetic activity. Neuroendocrinology 2005 May 26;81(1):63-68. [doi: 10.1159/000084894] [Medline: 15809515]
15. Kovatchev BP, Farhy LS, Cao H, Griffin MP, Lake DE, Moorman JR. Sample asymmetry analysis of heart rate characteristics with application to neonatal sepsis and systemic inflammatory response syndrome. Pediatr Res 2003 Dec;54(6):892-898. [doi: 10.1203/01.pdr.0000088074.97781.4f]
16. Ahmad S, Ramsay T, Huebsch L, Flanagan S, McDiarmid S, Batkin I, et al. Continuous multi-parameter heart rate variability analysis heralds onset of sepsis in adults. PLoS One 2009 Aug 14;4(8):e6642 [FREE Full text] [doi: 10.1371/journal.pone.0006642] [Medline: 19680545]
17. Monitor your heart rate with Apple Watch. Apple. URL: https://support.apple.com/en-us/HT204666 [accessed 2020-10-22]
18. Glasser JW, Hupert N, McCauley MM, Hatchett R. Modeling and public health emergency responses: lessons from SARS. Epidemics 2011 Mar;3(1):32-37 [FREE Full text] [doi: 10.1016/j.epidem.2011.01.001] [Medline: 21420657]
19. Swanson KC, Altare C, Wesseh CS, Nyenswah T, Ahmed T, Eyal N, et al. Contact tracing performance during the Ebola epidemic in Liberia, 2014-2015. PLoS Negl Trop Dis 2018 Sep 12;12(9):e0006762 [FREE Full text] [doi: 10.1371/journal.pntd.0006762] [Medline: 30208032]
20. Kang M, Song T, Zhong H, Hou J, Wang J, Li J, et al. Contact tracing for imported case of Middle East respiratory syndrome, China, 2015. Emerg Infect Dis 2016 Sep;22(9):1644-1646 [FREE Full text] [doi: 10.3201/eid2209.152116] [Medline: 27532887]
21. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]
22. Hirten RP, Stanley S, Danieletto M, Borman Z, Grinspan A, Rao P, et al. Wearable devices are well accepted by patients in the study and management of inflammatory bowel disease: a survey study. Dig Dis Sci 2020 Jul 23. [doi: 10.1007/s10620-020-06493-y] [Medline: 32705439]
23. Radin JM, Wineinger NE, Topol EJ, Steinhubl SR. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study. Lancet Digit Health 2020 Feb;2(2):e85-e93. [doi: 10.1016/s2589-7500(19)30222-5]
24. Ahmad S, Tejuja A, Newman KD, Zarychanski R, Seely AJ. Clinical review: a review and analysis of heart rate variability and the diagnosis and prognosis of infection. Crit Care 2009;13(6):232 [FREE Full text] [doi: 10.1186/cc8132] [Medline: 20017889]

Abbreviations

ANS: autonomic nervous system
HRV: heart rate variability
IBI: interbeat interval
ISMMS: Icahn School of Medicine at Mount Sinai
MESOR: midline statistic of rhythm
PCR: polymerase chain reaction
REWL: reweighted least squares
SDNN: standard deviation of the interbeat interval of normal sinus beats

Edited by G Eysenbach; submitted 27.11.20; peer-reviewed by G Luo, R Kadri; comments to author 18.12.20; revised version received 14.01.21; accepted 29.01.21; published 22.02.21.

Please cite as:
Hirten RP, Danieletto M, Tomalin L, Choi KH, Zweig M, Golden E, Kaur S, Helmus D, Biello A, Pyzik R, Charney A, Miotto R, Glicksberg BS, Levin M, Nabeel I, Aberg J, Reich D, Charney D, Bottinger EP, Keefer L, Suarez-Farinas M, Nadkarni GN, Fayad ZA
Use of Physiological Data From a Wearable Device to Identify SARS-CoV-2 Infection and Symptoms and Predict COVID-19 Diagnosis: Observational Study
J Med Internet Res 2021;23(2):e26107
URL: https://www.jmir.org/2021/2/e26107
doi:10.2196/26107
PMID:33529156

Original Paper

Understanding Gaps Between Daily Living and Clinical Settings in Chronic Disease Management: Qualitative Study

Mustafa Ozkaynak[1], PhD; Rupa Valdez[2], PhD; Katia Hannah[1], MPH; Gina Woodhouse[3], BSc; Patrick Klem[3], PharmD

[1]College of Nursing, University of Colorado | Anschutz Medical Campus, Aurora, CO, United States

[2]Department of Public Health Sciences, University of Virginia, Charlottesville, VA, United States

[3]University of Colorado Hospital, Aurora, CO, United States

Corresponding Author:
Mustafa Ozkaynak, PhD
College of Nursing
University of Colorado | Anschutz Medical Campus
Campus Box 288-18 Education 2 North Building
13120 E. 19th Avenue Room 4121
Aurora, CO, 80045
United States
Phone: 1 303 724 8273
Fax: 1 303 724 8559
Email: mustafa.ozkaynak@cuanschutz.edu

Abstract

Background: Management of chronic conditions entails numerous activities in both clinical and daily living settings. Activities across these settings interact, creating a high potential for a gap to occur if there is an inconsistency or disconnect between controlled clinical settings and complex daily living environments.

Objective: The aim of this study is to characterize gaps (from the patient's perspective) between health-related activities across home-based and clinical settings using anticoagulation treatment as an example. The causes, consequences, and mitigation strategies (reported by patients) were identified to understand these gaps. We conceptualized gaps as latent phenomena (ie, a break in continuity).

Methods: Patients (n=39) and providers (n=4) from the anticoagulation clinic of an urban, western mountain health care system were recruited. Data were collected through primary interviews with patients, patient journaling with tablet computers, exit interviews with patients, and provider interviews. Data were analyzed qualitatively using a theory-driven approach and framework method of analysis.

Results: The causes of gaps included clinician recommendations not fitting into patients' daily routines, recommendations not fitting into patients' living contexts, and information not transferred across settings. The consequences of these gaps included increased cognitive and physical workload on the patient, poor patient satisfaction, and compromised adherence to the therapy plan. We identified resources and strategies used to overcome these consequences as patient-generated strategies, routines, collaborative management, social environment, and tools and technologies.

Conclusions: Understanding gaps, their consequences, and mitigating strategies can lead to the development of interventions that help narrow these gaps. Such interventions could take the form of collaborative health information technologies, novel patient and clinician education initiatives, and programs that strongly integrate health systems and community resources. Current technologies are insufficient to narrow the gaps between clinical and daily living settings due to the limited number and types of routines that are tracked.

(J Med Internet Res 2021;23(2):e17590) doi:10.2196/17590

KEYWORDS

health information systems; workflow; self-management; activities of daily living; mobile phone

Introduction

Background

For patients with chronic conditions, health management is a continuous effort [1]. Management of chronic conditions entails numerous activities in both clinical (eg, developing plans of care) and daily living settings (eg, assessing medication side effects, controlling food intake) [2]. Activities across these settings interact [3-5], creating a high potential for a *gap* to occur if there is an inconsistency or disconnect between controlled clinical settings and complex daily living environments. We conceptualized *gap* as *a break in continuity in the performance of health-related activities across clinical and daily living settings*. A gap between health-related activities across settings is a latent phenomenon that can be identified or addressed only after precursors have surfaced or effects have been manifested (eg, when suboptimal health outcomes have occurred). As an example of a gap [4], medication prescription is a part of the clinical workflow that requires coordination with the patient's daily living environment. Not incorporating daily living information to tailor therapy could lead to nonadherence. The study by Ozkaynak et al [4] provides more information on the definition of a gap. Identifying the specific points where a health care regimen can or must be modified is possible only if all health-related activities in both settings are considered part of a health care continuum [4-8]. Adding a gap construct to known models and frameworks may allow for a better comparison between theoretical and real-world management systems for patients.

This study aims to understand why these gaps occurred, what effects they produced, and preventative or remedial actions. The work was guided by the *Infinicare* framework [4], a theoretical perspective that highlights the need for integration of health-related activities across care delivery settings and accounts for patient-specific contextual features. The *Infinicare* framework has 3 principles. First, clinic-based health-related activities shape or influence activities in the context of daily living. Second, health-related activities within daily life inform activities in the context of the clinical setting [4]. Third, the *Infinicare* framework highlights the importance of context and operationalizes it by incorporating 4 different contextual dimensions: physical, organizational, social, and cultural. These 4 dimensions of context interdependently affect health-related activities in both clinical and daily living settings. The *Infinicare* framework posits that health and health care have no strict temporal or physical boundaries.

Literature Review

Connecting Fragmented Health Care Services for Continuity of Care

Traditional studies of fragmentation within health care delivery and continuity of care focus exclusively on the clinical setting and demonstrate the adverse consequences that result from a failure either to defragment or to establish continuity of health care services [9,10]. Interventions (eg, secure messaging, mHealth, communication skill building) designed to address gaps across clinical settings have improved outcomes [11-14]. Our work seeks to expand this paradigm by understanding latent gaps through precursors and consequences that occur *between* clinical and daily living settings. Approaches such as patient-oriented workflow [15], patient journey [16], and adherence work [17] posit that health-related activities in the clinical and daily living settings are parts of a broader workflow. Research connecting clinical and home settings has focused on postacute care as related to hospital discharge. Gaps that occurred after hospital discharge (for both acute and chronic conditions) have focused on adverse drug events, avoidable hospitalizations, and medical errors [18,19]. A limited number of studies have examined the interplay between clinical care and self-management in the daily living environment [3,17,20]. Identifying factors that contribute to gaps [18] can inform interventions such as standardized communication protocols [21], secure messaging [22], personal health records [1], medication reconciliation strategies [23], patient and family education [24], community-based support [24], and patient navigators [25].

Interventions for continuity of care across settings have been institution centric [26]. These interventions (eg, discharge planning, home-visiting programs) are not necessarily tailored to an individual's unique daily living circumstances, which obviates a thorough understanding of how these interventions may affect therapy and patient outcomes. Moreover, being institution centric, these interventions focus on the clinician rather than emphasizing patient-clinician collaboration as the driver of optimal outcomes [27,28]. Models and frameworks to describe daily living–related barriers to adherence have been developed (eg, patient work system [2]).

The study of gaps across settings is more challenging, as it requires the application of methods for studying continuity of care in clinical to daily living settings, the latter having more variability [5,29]. Variabilities include individual characteristics (eg, preferences, values), physical environment, social environment, cultural factors, and organizational (eg, community) resources. A study of older adult patients with heart failure applied a *work system* to patient-performed work in self-management. This study highlighted the interaction between systemic components (eg, tasks, technology) in which the patient was embedded [2]. However, more research is needed to examine differences in how the patient work system is conceptualized to pinpoint gaps, the consequences of gaps, and strategies for addressing these areas.

Collaborative Health Information Technologies

We define collaborative health information technology (HIT) as an information technology (IT) that has a full functional interface with clinicians, patients, and all other care partners. Collaborative HIT refers to a combination of connected technologies (eg, personal health records, electronic health records [EHRs], mHealth apps, social media) or an ecosystem [30]. The use of collaborative HIT can potentially provide business value, socio-organizational value, technical process, and evolution value [31].

Collaborative HIT [32] is underdeveloped in terms of facilitating education, integrating resources, and allowing for information exchange [30]. For example, provider access to patient-generated

data and patient access to personal health information are limited. This study is timely in that isolating and capturing gaps in health care delivery will inform the design of collaborative HIT for use by both patients and clinicians.

Most informatics tool-design models begin with a needs assessment [33,34]. To our knowledge, previous systematic, large-scale, qualitative needs assessments for health IT design have focused on either clinical activities or supporting health-related activities in daily life, not the space or gaps between these areas [35,36]. Consequently, the purpose of this study is to characterize gaps between clinical and health-related activities in daily life, using anticoagulation treatment as an example. This study is timely given that collaborative HIT is in its very early stages of development [30].

The contributions of this study include the following: a systematic exploration of reasons for gaps, consequences of gaps between diverse settings, strategies for stopping these gaps via a systems framework, and a patient-oriented workflow approach. Study findings have the potential to influence how existing interventions could be reimagined or expanded through design recommendations for collaborative health IT [37-39].

Case Exemplar

Anticoagulation therapy with warfarin is a particularly rich example for understanding gaps between clinics and daily life [7]. It requires strict adherence by patients and collaboration with providers, as frequent dosing adjustments are often necessary [40]. Food and alcohol affect dosing; therefore, a strict limitation of foods high in vitamin K and limiting alcohol intake is required. Moreover, contextual factors such as family support, ability to purchase food, and avoidance of specific foods that can affect therapy must be considered. Monitoring the therapeutic dose of warfarin is achieved by measuring blood levels for clotting. The international normalized ratio (INR) is a measure that evaluates the therapeutic range of anticoagulation [41]. Patients require testing in the clinic setting as often as several times a week, as home anticoagulant monitoring was not available at the site studied. Warfarin or Coumadin interacts with common antibiotics; therefore, communication is essential by the patient with their non–clinic-based providers (eg, dentist, podiatrist) ordering an antibiotic. Good communication between the provider and patient about testing results and dosing decisions is crucial. A wrong dose can result in serious bleeding, thromboembolic events, and medical complications from the underlying medical problem for which the anticoagulant was ordered. Examining gaps in optimal treatment can reveal clinician, patient, and system-related factors leading to suboptimal dosing and inform interventions.

We recognize that the dynamics of other chronic conditions (eg, diabetes, HIV) may differ from anticoagulation management. The findings and design guidelines based on anticoagulation, however, have the potential to be translated where HIT supports self-management activities in individualized therapy plans across patient conditions.

Methods

Study Design

We used a qualitative study design to examine and characterize gaps between health-related activities across diverse settings. Qualitative design allows for exploring and gathering rich descriptions of the phenomena of interest (ie, gap) in a context. Data were collected from patients and providers (clinical pharmacists) in a single academic, hospital-based anticoagulation clinic through interviews and journaling. Data were analyzed qualitatively. This study was approved by the Colorado Multiple Institutional Review Board.

The *Infinicare* framework [4] shaped this study in 2 ways: (1) it guided the development of the data collection tools by ensuring that data were collected on basic components (eg, health-related activities in daily living and clinical settings and social, organizational, physical, and cultural context); and (2) guided data analysis through use of select *Infinicare* constructs (eg, daily routines, challenges of self-management, clinicians' understanding of the patient's daily living environment) for the initial codes.

Settings and Sample

Patients were recruited from the anticoagulation clinic of an urban, western mountain health system in the United States. The clinic provides outpatient care and is staffed by 4 clinical pharmacists. The primary medication used by the clinic was warfarin, an anticoagulant in pill form. Most clinic visits were face-to-face; however, phone visits also took place.

Exclusion criteria were individuals younger than 20 years of age. In an effort to approximate a real-world setting, participants with comorbid medical conditions, mental health conditions, substance use disorders, or pregnancy were eligible for inclusion. We used a stratified sampling strategy to ensure that we recruited older adult patients as well as younger patients of both genders and those both new and experienced with anticoagulation therapy. One participant's first language was Spanish, for which a bilingual translator was used for data collection.

Recruitment

Fliers were posted in the clinic with information about the objective of the study, sponsors, principal investigator, institutional review board approval number, and contact information for the research team. Interested participants called the dedicated phone number to provide their contact information. This information was accessed only by the research team, who called to verify that the prospective participant met the inclusion criteria.

Data Collection

Data were collected (Textbox 1) first from primary interviews with patients, followed by patient journaling on tablet computers and exit interviews. We conducted provider interviews as the last step in data collection. Data from exit interviews with patients and provider interviews were collected to obtain both views. Figure 1 summarizes the data collection methods and how the different methods related to each other. Complete data

(primary interviews, journaling, and exit interviews) were obtained from 90% (35/39) of patients (4 completed only the primary interviews). Four providers participated in the in-person interviews. Each patient interview (primary and exit) was conducted in person by the first (MO) and third authors (KH). These interviews occurred in a private meeting room in the College of Nursing, which is in the same campus as the anticoagulation clinic.

Textbox 1. Steps in data collection and information elicited.

Step 1: Primary interviews (patient data collected)

- Care received by provider
- Social & physical environment
- Health-related activities
- Tools and technologies used
- Challenges and facilitators carrying out activities
- Interaction between clinical and daily activities

Step 2: Journaling (patient data collected)

- Health-related activities
- People involved
- Social, cultural, and physical contexts in home

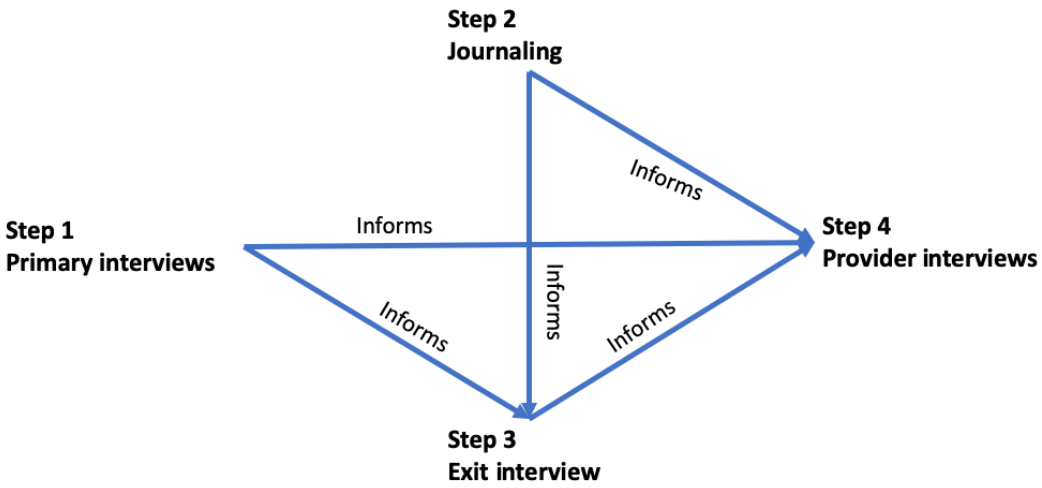
Step 3: Exit interviews (patient data collected)

- Further exploration of issues discussed in Steps 1 and 2

Step 4: Provider interviews (provider data collected)

- Evaluation of health-related activities
- Description of clinical processes, policies, and development of therapy plan
- Provider

Figure 1. Data collection tools.



Those conducting the interviews used a predetermined list of questions; however, participants were also asked follow-up questions for clarification and to gain additional insights. *Primary patient interviews* provided information related to: (1) the care they received in the anticoagulation clinic; (2) their social, cultural, organizational, and physical environment; (3) daily health-related activities; (4) tools and technology they used to support these activities; (5) challenges and facilitators while conducting these activities; and (6) manner in which clinic-based health-related activities shape or influence activities in daily living contexts and health-related activities within daily life inform activities in the clinical context. *Journaling* provided

data about health-related activities, the people involved in the activities, the temporal organization of the activities at home and other daily living settings, and the social, organizational, cultural, and physical contexts in which the activities occurred. Journaling also partially mitigated the recall bias. Journaling was accomplished by a tablet application that made voice entry possible. The application also moved all audio recordings to secure servers to facilitate data collection and analysis. Patients were provided approximately 20 min of training on the use of the tablet and the app. During the training, participants were instructed on how the journaling app worked, and the researcher reviewed the 11 guiding questions with participants. Participants were asked to journal at least every 2 days for a month and provided the telephone number of the research line in the event they needed assistance. Patients were given an option to keep the tablet computer as a thank for participating. *Exit interviews* were conducted with the patients after they completed journaling and explored issues that had been mentioned in the primary interview and journals. Exit interviews (Step 3) were conducted in person in the same room as primary interviews, with information used for triangulation. Patient primary and exit interviews averaged 48 and 20 min, respectively. *Provider (pharmacist) interviews* (n=4) lasted approximately 73 min. During this time, health-related daily living activities reported by the patients were explored and a description of the clinical processes, clinical policies, development of therapy plans, and provider perspectives dealing with challenges related to adherence of the patient.

We received consent from 39 patients for the primary patient interviews, journaling, and exit interviews and received consent from the pharmacists for provider interviews. For all of the data collection, any reference to a personal identifier (eg, names of people, organization, etc) was removed.

Data Analysis

Interviews were transcribed verbatim professionally. The journal entries were reviewed and transcribed by KH. Qualitative analysis was based on a theory-driven approach [42] and the framework method of analysis by Gale et al [43]. The main concepts of the *Infinicare* framework [4] provided *a priori* codes, with additional codes identified inductively when textual elements did not fit within these predetermined codes [44,45]. The advantage of using an overarching framework was to increase the efficiency of the data analysis process and to build on prior studies. The advantage of deriving additional codes based on collected data allowed for better integration of the participant's experiences.

Data from the 4 collection methods were coded, main themes were identified, and relationships between the main themes were examined. Data were managed using DEDOOSE 5.1.18. The analysis included independent coding of approximately 10% of the data by MO and KH. During this coding, various code ideas were discussed in light of the broader study purpose, such as (1) the process and contextual components associated with care for anticoagulation management and (2) the potential disconnect between health-related activities in the clinical setting and outside the clinical setting. Coding was then reviewed by 3 team members (MO, KH, and RV) until consensus was reached, following which a codebook was developed [46]. The codebook included codes, the concepts that were represented by the codes, specifications (definition, inclusion, and exclusions as applicable), and examples. We also allowed simultaneous coding, meaning that multiple codes could be applied to a piece of text [47]. KH applied the codebook to the remainder of the data. MO then examined embedding the codes within the overarching categories, based on their content.

Results

Description of the Participants

Study participants ranged in age from 25 to 83 years and had been receiving anticoagulation therapy for 2 weeks to 26 years (Table 1). Most participants were women and White. Educational attainment varied, although most had a minimum education of a college or associate degree (30/39, 76%). Most participants lived with someone. The median household income was US $60,000; 3 of the 4 providers had a PharmD degree, the other had a BS degree. The average experience of the providers was 15 year.

The gaps are the operationalizable aspect of the *Infinicare* framework and can be more directly used to guide HIT development. Our analysis revealed 3 aspects of gaps: reasons, consequences, and strategies to overcome these gaps (Textbox 2).

Table 1. Patient participant demographics from anticoagulation clinic (N=39).

Demographics	Values
Age (years)	
Median	56
Range	25-83
Mean (SD)	53 (16)
Therapy duration	
Median (years)	3
Range	2 weeks to 26 years
Gender, n (%)	
Female	23 (59)
Ethnicity, n (%)	
White	27 (69)
African American	9 (23)
Highest education degree, n (%)	
Masters	7 (18)
College	8 (21)
Some college or associate degree	15 (38)
High school	5 (13)
Less than high school	4 (10)
Living situation, n (%)	
Alone	6 (15)
Spouse only	12 (31)
Spouse and ≥1 child	7 (18)
Family members or parents	9 (23)
Roommates	4 (10)
Other	1 (3)
Income (US $)	
Median	60,000

Textbox 2. Reasons, consequences, and strategies used to overcome gaps.

Reasons for Gaps

- Recommendations did not fit into patient's daily routine
- Recommendations did not fit living context
- Information not transferred across settings

Consequences

- Cognitive and physical workload on the patient
- Poor patient satisfaction
- Compromised adherence to therapy plan

Strategies

- Patient-generated strategies
- Routines
- Collaborative therapy management
- Social environment
- Tools and technologies

Reasons for Gaps

The reasons for gaps were from 3 areas related to therapy recommendations: misfit with the patient's daily routine, misfit within the context of daily living, and information was not transferred across settings.

Recommendations Did Not Fit Into Patients' Daily Routine

Daily routine is a structured, temporal cycle of activities that occurs daily or in some instances weekly. Participants noted difficulty integrating new activities (eg, cognitive remembering to take medication and physical-going to appointments) into their existing personal routines. These new activities competed with other activities in terms of time, money, and attention (Table 2, Quote ID #1 [ID 1]). Having a busy schedule (multiple jobs and family-related responsibilities) further worsened the challenge and required devoting additional time to planning the day. This was particularly emphasized by participants as an issue in planning meals (ie, deciding what to cook or shop for) and monitoring food consumption. Participants reported changes from previous meal-planning routines or grocery shopping norms. However, new routines may not be integrated and adapted quickly. Those with comorbidities requiring multiple therapies have magnified this challenge. In the words of one participant, additional therapy-related activities to an already busy schedule "slows [them] down." Activities necessitated by therapy recommendations only fit into strict daily routines. Weekends, travel, seasonal, or religious events made adherence difficult (ID 2).

Therapy recommendations were not easily integrated into social and extracurricular routines. Food choices influence the effectiveness of the anticoagulant, requiring changes to dietary habits not only for participants but also for their families. Alcohol consumption has a strong influence on dosing. As alcohol also serves as a social activity, therapy recommendations necessitated changes in social routines such as avoiding social situations where drinking occurred. Participants also avoided social engagements to evade potential questions about their health (ID 3). Finally, due to the potential for serious bleeding from trauma (a side effect of anticoagulation), participants reported relinquishing some hobbies (eg, mountain biking; ID 4) or work-related activities (eg, riding horses).

Table 2. Quotations supporting selection of categories and subcategories.

Category, subcategories, and quote ID	Sample quotes	Source of quotes
Reasons for gaps		
Therapy recommendations do not fit into daily routine (additional cognitive and physical steps; busy life schedule is a challenge; planning challenges; comorbidity; variation in routines; decreased social engagement; giving up some hobbies)		
1	"It's nearly 8 o'clock and I'm just now getting home I left the house at 6:30 this morning. I didn't pack a lunch and who knows what I ate in terms of Vitamin K and now I have to go back to remember try to figure out what I'm having for dinner, and I threw up because I'm still sick so that'll be fun to try to guess again today" (additional cognitive and physical steps; busy life schedule is a challenge; comorbidity)	Journaling
2	"I forgot to take my medication this morning at the normal time. Since I have the day off and the next day off with the holiday, I am on a little different schedule so I am going to have to be a little more conscious that I don't start my day without taking my medication like I did today." (planning changes; variation in routines)	Journaling
3	"Went to a new neighbors house they didn't realize that I couldn't drink. [I] didn't want to explain it so just said that I was designated driver even though we were all just walking home, but I hate having to explain" (decreased social engagement)	Journaling
4	"... knowing I'm on Warfarin...they don't want you to get lightheaded or anything like that. So I can't jog like I want to" (giving up some hobbies)	Primary patient interview
Therapy recommendations do not fit into patients' living contexts (social; organizational; physical; cultural; temporal)		
5	"My husband wants to have a variety in the diet. I really can't do that. I have to keep my vitamin K levels consistent. That's the reason I tend to eat a lot of the same foods." (social)	Journaling
6	"Sometimes I get frustrated because I need afternoon appointments, period. I need afternoon appointments and I have to try to get them on Thursdays as much as possible, because that is the only time transportation is available." (organizational)	Primary patient interview
7	"It is one bedroom, so you don't get any privacy and that is a little bit difficult." (physical)	Primary patient interview
8	"I know my highest INR[a] was 3.9 and it was after my stepdad passed away I hadn't been eating salads. I was back in Michigan for 10 days. Hadn't really been eating healthy. And definitely taking in more alcohol. That was kind of a scary thing. When you know that, it is wow! The possibility of if you had an accident and bleeding more would be greater." (social)	Primary patient interview
9	"There's been so much going on because of Christmas and New Years and my son's birthday. Which was hard for me, because with Christmas and shopping and everything else that was going on during the last month, it was a little difficult to stay on task." (cultural; temporal)	Patient exit interview
Necessary information is not transferred across settings (insufficient transfer from the daily living to clinical settings; insufficient transfer from clinical settings to the daily living)		
10	"I usually lie to [providers]. If they say have you missed a dose I say no. Cause it is just easier than saying well, actually 3 weeks ago I did. Or ...I figure by then it doesn't really make a difference. I'd probably tell them if it was within 48 hours that I missed a dose and my bloodwork was off. Then I'd probably confess to it." (insufficient transfer from the daily living to clinical settings)	Primary patient interview
11	"A lot of challenges are misconceptions. Often patients will say, I can't take this. I don't like taking Warfarin because I can't do activities. I'm afraid I'm going to fall. I'm going to hit my head and have a fatal bleed. Whereas in reality, with Warfarin, you shouldn't let the medicine control your lifestyle... But we tell people to be safe." (insufficient transfer from clinical settings to the daily living)	Provider interview
Consequences		
Additional cognitive and physical patient workload (remembering; decision making; physical workload)		
12	"the biggest one is probably just remembering to take [The medication]. Before being diagnosed with this, I never had anything that I had to take every single day." (remembering)	Primary patient interview
13	"I was having problems managing taking my...Antibiotics! 2 hours before taking my other meds. My anticoag. the decision was either I skip the antibiotics, or I skip the anticoag. So I skipped the antibiotics. But that seemed to happen like 4 times this last month." (decision making)	Patient exit interview
14	"And that is still a struggle with just being busy. And then also, in that, I have to kind of watch. If I know I'm going to be out of town, then I need to make sure not to run out of my prescription. To make sure I get things refilled and just being more strategic with the pharmacy and making sure I have refills and that I get there before I leave, if I'm going somewhere or something like that." (physical workload)	Primary patient interview

Category, subcategories, and quote ID	Sample quotes	Source of quotes
Poor patient satisfaction (difficulties and inconvenience; frustration; overwhelmed)		
15	"The best is definitely with the menstrual cycle. That is just something that is in my family. And it is just very, very heavy within our family. Heavy cycles. And if I continue to take the Coumadin during that time, for like 2 of the days, I literally cannot leave the house. It is just too heavy. So I have to stop during that time. And the previous providers just did not understand that." (difficulties and inconvenience; frustration)	Primary patient interview
16	"It is getting better, but it is almost more of an annoyance at this point. That it is always there. That it is always part of what I have to think about. But it is just kind of become like your multiplication facts. You've memorized it. It's there. It is just part of memory space, at this point." (frustration; overwhelmed)	Patient exit interview
Compromised adherence to therapy plan (unintentional; by choice)		
17	"I forgot to fill my weekly pill tray. And so tonight I took the wrong dose and so I need to fill it up with the right dose." (unintentional)	Journaling
18	"Accepting the fact that sometimes I try to sneak like cranberry juice or grapefruit, like when no one is paying attention." (by choice)	Primary patient interview
Strategies		
Patient-generated strategies (self-adjusting the dose; green-alcohol balancing; timing of medication; remembering mechanisms)		
19	"After dealing with [bleeding nose] for years I finally said I'm not doing it this way anymore and I started changing my own dosing, according to my symptoms." (self-adjusting the dose)	Patient exit interview
20	"And when I was drinking, I was drinking, you know, quite a bit. Like daily almost. (laughing) And yeah, I was real cognizant of making sure I took in even more greens to counteract the alcohol" (green-alcohol balancing)	Primary patient interview
21	"Why 'The Price is Right?' Because it is something that helps me...I just woke up one day and I took it and the 'Price Is Right' was on. And then the next day I saw, oh well, the Price is Right is on, it's time to take it." (timing of medication; remembering mechanisms)	Primary patient interview
Routines (medication use; food consumption)		
22	"I take my Coumadin usually at 4:30 every afternoon. And that's when we feed the dogs. So it is kind of a routine. And when we are on the road, we do the same thing. The dogs have to be fed at 4:30. There ain't no ifs, ands or buts." (medication use)	Primary patient interview
23	"I guess [providers] got it to be my routine. [providers] figured out what I need. And I tell them ok. I eat Pizza on Wed. Thurs. I'm going to have my Spinach. Friday I'm going to have chicken. Then I make soup. And you know? And then I pretty much...And I got on a new thing where I make Shrimp Louie. I try to have Shrimp Louie every week or two. And you know...So I guess my diet doesn't change a whole lot. Old Irish guy. Meat and potatoes. So that is kind of where I think it makes it easy for me. I guess I'm kind of mundane and routine." (food consumption)	Primary patient interview
Collaborative therapy management (patients being involved; plans around patients' life; shared decision making; establishing trust)		
24	"Every time I have questions about something like, when winter is coming and I know I am going to be eating less fresh greens or something, [clinicians] are willing to say ok, well we can proactively start to cut back your Warfarin if you think you are going to be eating less and then see how you are doing." (patients being involved; plans around patients' life; shared decision making)	Patient exit interview
25	"I write down a note to myself like, typically I would say, [this patient] eats greens two to three times a week. But his INR was low because he had collard greens over Thanksgiving and that made his INR low. So I know I'll give him a little extra and then we'll go back to what he was doing before." (plans around patients' life)	Provider interview
26	"So once you start to get to know [patients], you kind of get to know how aggressive you can be with making adjustments and how conservative you should be." (plans around patients' life)	Provider interview
27	"In terms of my relationship with [the provider] as far as trust goes. I trust her recommendation when she does or does not adjust my dosage. It makes me feel more comfortable with her to ask questions. I think it makes a huge difference. And also in my willingness to keep coming back." (patients being involved; establishing trust)	Primary patient interview
Social environment (emotional support; experience sharing; protecting risks; make suggestions; transportation; reminder; prepared food)		

Category, subcategories, and quote ID	Sample quotes	Source of quotes
28	"I got an ear full of people. My girl, my mom, my sister. So support is there, to the point of where I'm going to do what I have to do. Cause I'm not going to put up with these people. So it's there and they are real concerned and making sure I'm taking care of myself and on my medication, like I'm supposed to. I got a very good support base" (emotional support)	Primary patient interview
29	"My mom has been on it since she was 16, so I've learned so much from my mom." (experience sharing)	Patient exit interview
30	"Collaterally, they [friends] help. Like if I need to do some chores or heavy lifting, or something that puts me at risk for getting bruised or beat up, and they are there, then they will usually jump in and give me a hand." (protecting risks)	Patient exit interview
31	"The lady at my job, who had the stroke from the blood clot - she is still alive. She told me to do Yoga. She said Yoga is a good substitution that will help replace jogging." (make suggestions)	Primary patient interview
32	"I try to schedule on Monday's now so that my fiancé' can drive me, cause he doesn't work on Monday's now. So he will drive me and then it's not that big of a deal." (transportation)	Primary patient interview
33	"My wife is always reminding me probably about every night. So she stays on top of it." (reminder)	Patient exit interview
34	"[roommates] contribute a great deal, they cook certain things to make sure I can eat it." (prepared food)	Journaling
Tools and technologies (tools that patients bring in; tools that clinic provides; INR machines)		
35	"Really the only challenge is remembering medicine and I use a pill container to get over that challenge." (Tools that patients bring in)	Journaling
36	"[Providers] give me calendars where I can write it down. And I cross it off when I take it. And when I don't take it, they know when it's blank I didn't take the medication." (tools that clinic provides)	Patient exit interview
37	I'd go back to being able to do the INR at home. (INR machines)	Primary patient interview

[a]INR: international normalized ratio.

Recommendations Did Not Fit Into Patients' Daily Living Contexts

We distinguished the living context from daily routines. Living context is a manifestation of the characteristics of settings that serve as obstacles or facilitators to performing routines. For example, the social context highlights pressure and encouragement from others rather than the fact that having a routine that involves a social life that is inherently challenging.

Participants reported misfits of anticoagulant therapy within their social, organizational, physical, cultural, and temporal contexts. Their social environment could adversely influence their consumption of foods, resulting in conflict with their therapy (ID 5). Misfits related to the organizational context included lack of access to transportation to the clinic and limited clinic hours and their work schedule (ID-6). Challenges related to the physical context related to limited residential space, resulting in a lack of privacy and need to spend more time outside or away from home (ID 7). Clinicians reported that cultural activities such as fasting practices for religious purposes and excess consumption of green tea could be problematic. Participants reported that the temporal context (eg, Thanksgiving, birthdays) came with new requirements that made therapy management difficult (ID 8 and 9).

Information Not Transferred Across Settings

Insufficient transfer of information from daily living to clinical settings was represented by incomplete or inaccurate patient history (ID 10). This occurred through misconception by patients regarding therapy, not recalling the therapy plan, and/or misinterpreting the provider's comments. Rushed encounters, patient misconceptions, and organizational policies (such as not accepting patients late for an appointment or lack of available timely appointments) could lead to insufficient transfer of information from clinical to daily living settings (ID 11). Missing an appointment means there is no transfer of information in either direction, resulting in an inability to create a plan designed specifically to suit the patient's lifestyle.

Consequences of Gaps

We grouped the consequences of gaps into 3 themes: cognitive and physical workload on the patient, poor patient satisfaction, and compromised adherence to the therapy plan.

Cognitive and Physical Workload

Cognitive workload issues included additional remembering and decision making. Examples include the type and quantity of recent food consumption, taking medication at the correct time and dosage, obtaining medications from the pharmacy, and making appointments (ID 12).

Additional decision-making tasks were decisions during grocery shopping, cooking, plate portions, and ordering at restaurants to ensure that vitamin K consumption was consistent. Almost all participants took warfarin daily, but the daily dosage could vary with the day of the week. On occasion, a decision had to be made on whether to skip a dose if the medication was not taken as scheduled (ID 13). Participants faced implicit decisions

about whether to focus on adhering to therapy versus engaging in activities that may potentially jeopardize patient outcomes. Physical workload issues included challenges related to attending appointments, obtaining medications from a pharmacy, and traveling back home due to forgetting medication (ID 14).

Analysis highlighted that some causal links (anecdotal level) between the reasons and consequences of gaps are related. ID 12 exemplifies the additional cognitive workload on patients, when additional steps are not well integrated into daily living routines, and possibly the need for preplanning. Busy lifestyles exacerbate this misfit.

Poor Patient Satisfaction

Increased workload is a phenomenon in which patients have an emotional response that can affect health-related quality of life. This emotional response falls into the theme of satisfaction. For example, dealing with bleeding was reported as a burden on physical workload; however, the emotions of frustration, fear, and shame (as it relates to stigma) would fit well under satisfaction. Participants reported a wide range of negative emotions due to the challenging experiences resulting from the gaps reported above. One participant reported excessive bleeding during her menstrual cycle, making it difficult for her to leave home, which led to frustration and inconvenience (ID 15).

Other reported frustrations included food constraints, the need for food tracking or remembering, and renewing prescriptions. Some participants reported a social stigma and embarrassment: "talking about anticoagulation makes [them] feel old."

The therapy requirements could be overwhelming. The therapy could be an annoyance since "It is always a part of [the participant's] life."

Compromised Adherence to Therapy Plan

Adherence was not a problem for all patients; however, reduced adherence could lead to serious health consequences. Participants reported the following adherence related issues: delay in taking medication, taking a wrong dose, missing a dose, and not following the dietary restrictions (ID 17).

Some instances of lack of adherence were by choice. Some reported "sneaking in" contraindicated foods (ID 18), consuming excess alcohol at special events, and missing medication when away from home. Adherence could be due to misfit of the therapy plan with the patient's daily routines (ID 2) or the context of their daily living.

Dealing With Gaps

Five strategies were used to cope with the challenges of gaps: (1) patient-generated strategies, (2) routines, (3) collaborative therapy management, (4) social environment, and (5) tools and technologies.

Patient-Generated Strategies

We defined patient-generated strategies as improvisations (not necessarily approved by their clinician) by patients with the intention of better self-management. These strategies included avoiding adverse events (bleeding) and helping with remembering. Strategies were typically discovered by patients over time and by *trial and error*.

One common strategy was self-adjusting the medication dose to avoid bleeding and symptoms (ID 15 and 19). Some reported the consumption of greens when they had consumed alcohol, under the assumption that greens balanced out the effects of alcohol (ID 20). Other reported patient-generated strategies included improvising remembering mechanisms, training the dog to bring the phone, and carrying extra medications (ID 21).

These strategies were not necessarily suggested by clinicians, but clinicians were often aware of their use. Clinicians reported mixed opinions about the effectiveness or harm of these strategies.

Routines

Routines were defined as repeating the same set of activities at approximately the same time of the day. Participants utilized routines for cognitive support (mostly remembering). For example, medication use was timed by daily events (feeding dogs, a TV show, and taking a shower) or a specific time of day (ID 22). They also used weekly events or times to fill their pillbox. Grouping and placing medications in different areas of the home was another way of routinizing medication use. Routines regarding food consumption included frequenting the same restaurant that provided the same food and having a fixed weekly menu (ID 23).

The majority suggested that establishing routines helped with therapy management; however, busy periods made the following routines more difficult. New routines were easier to follow if they fit into their existing daily routines. Participants reported breaks in routine during holidays.

Clinicians reported that routines were one of the main strategies they suggested; however, the risk of establishing routines was difficult to *alter* with changes in therapy. Summarized by one pharmacist as, "habits are hard to form and hard to break."

Collaborative Therapy Management

Collaborative therapy management (CTM) was defined as active efforts by patients and clinicians (and health care system with the available resources and opportunities to the patient) to work together for optimal health outcomes. Participants contributed to CTM by providing an accurate history, being involved in decision making and engaging actively in their health (eg, prioritizing health-related activities and taking responsibilities, ID 24). Clinicians contributed to CTM by asking personalized questions, developing therapy plans around the patient's life, making it easy to comprehend and apply, encouraging engagement, and providing tools that allow patients to take responsibility for their health (ID 25).

Therapy plans required knowing the patient's lifestyle, individualizing recommendations congruent with the preferences and routines of the patient, and adjusting dosages around the patient's daily life (ID 26). Making the therapy easy included being flexible in appointment times, expanding office hours, having processes that allowed patients to use local laboratories, and accommodating transportation limitations.

CTM also included shared decision making and iteratively adjusting therapy until a solution that was mutually agreed upon was reached. Both clinicians and patients highlighted that

establishing a trusting relationship was a prerequisite for successful CTM (ID 27). CTM is particularly important for patients with complex needs due to multiple comorbidities and living situations. In particular, CTM highlighted difficulties in using the strategies described.

Social Environment

The patients' social environment varied and included: family (mother, sister, grandchildren, aunt, and in-laws), friends (social media as well as in person), colleagues, partners, pastors, and roommates. The social environment provided emotional support by not eating restricted food in front of the patient, looking out, being mindful, encouraging, sharing funny material through social media, showing understanding, and "checking-in" periodically (ID 28). The social environment goes beyond just providing emotional support but also logistical, cognitive, and informational support. Some found it helpful to have others convey their experience with the same therapy (ID 29), protection from risks (ID 30), made (food and exercise) suggestions (ID 31), provided transportation (to appointments and for medications; ID 32), reminded of medication use and food restrictions (ID 33), and prepared appropriate food (ID 34).

Tools and Technologies

The most commonly reported tools and technologies included a pillbox, reminder alarm on smartphones, and a calendar tool provided by the clinic (ID 35 and 36). This calendar displayed the days when the medications were received and skipped. Some reported their use of smartphone apps to track food consumption. Many reported their willingness to use INR machines at home; however, the clinic did not support it (ID 36).

These strategies help narrow the gap by reducing the effects of the 3 aforementioned reasons and mitigating the consequences. Routines and social environment can help with better fit of therapy plans, and CTM can mitigate patient dissatisfaction.

Discussion

Principal Findings

This study examined chronic disease–related challenges by focusing on gaps between therapy plan development (ie, clinic based) and implementation (ie, daily living based), using the case of anticoagulation therapy. Gaps were latent and not directly observable, but their effects were evident. We ensured the collection of rich data on gaps by employing multiple data collection methods that inform each other. We also collected data from both patients and providers. The use of a framework (ie, *Infinicare*) ensured that only relevant data were collected on all aspects of the gaps. The reasons for and consequences of gaps as well as strategies to overcome them were identified. The reasons for gaps can prompt needs assessments and inform systematic interventions (eg, informatics, policy, educational, operational) that narrow these gaps. The consequences should also be taken into account when designing and implementing interventions. Moreover, these consequences can serve as evaluation criteria for interventions. Interventions should be congruent with the 5 reported categories of strategies for bridging the gaps. The reported strategies are a jumping-off point, and other strategies may be developed.

Understanding the reasons for the consequences of and strategies to overcome gaps can also help integrate the concept of gap into the current and future conceptual models and frameworks of health management. These models and frameworks can better explain the factors that affect health management.

Previous studies [26,48] that focused on the individualization of health care services and the transition of care have highlighted some of the themes discussed in this paper, such as collaboration among clinicians, patients, and family members; social support; provider awareness and knowledge of patient situations; understanding patients' cultural practices and beliefs; understanding places in which the patient lives; emphasis on the patient and family experience; focus on home and community-based care; emotional support and alleviation of fear and anxiety; and encouragement of continued social roles. Moreover, studies that focused on self-management activities and adherence work stressed the importance of daily living contexts [2], discrepancies between clinicians' and patients' perspectives on self-management [49], and localization of therapy management [20]. However, this study went beyond previous findings by identifying the *concept of the gap*. Conceptually, gaps can provide more insights to researchers by better explaining incongruence among work system elements [2] in chronic disease management. This study provides a foundation for collaborative HIT design guidelines by highlighting relationships among the precursors to, consequences of, and counter measures against cross-setting gaps.

Design Implications

The 3 reasons identified for gaps underscored the need for clinician access to HIT that would incorporate more information about the patient's daily living environment. Current consumer informatics technologies (eg, personal trackers, food log applications) support data collection on only a limited number of activities of daily living (eg, exercise, food consumption); however, these data are not always leveraged in developing therapy plans taking into consideration a patient's routines and context [50,51]. This study shows that other types of routines (eg, social life) also need to be captured and accounted for. Moreover, there is a need to capture the interplay between different types of routines (eg, social life, food consumption, medication administration). The use of such rich data in creating therapy plans would require better integration of patient-generated data into EHRs. Clinical decision support systems that use machine learning algorithms can harness historical clinical and patient-generated data and assist in developing individualized therapy plans [52,53]. Mobile technologies provide opportunities to better capture essential information in each clinical and daily living setting and information that can then be retrieved as needed across settings [54-56]. We also contend that, instead of prioritizing either clinical or daily living settings, technologies should foster collaborative dialog. This collaborative dialog should include (1) joint clinician-patient interpretation of patient-generated data [39] and (2) discussion of patient-driven therapy plan translation options [20]. The reasons for gaps can guide

clinicians' question-asking and decision-making behavior. Focusing on gaps can help inform interventions with (1) technology, (2) policy, (3) educational, and (4) operational components that would support management in chronic disease. In fact, any collaborative HIT intervention that would narrow or close the gaps could need all 4 components. In addition, HIT designs can also be developed to promote sustained collaborative relationships between clinicians and patients [32,52]. There are no explicit guidelines on how to design and implement a truly collaborative HIT. This paper creates a starting point for the design and implementation of collaborative HIT. Design efforts should ensure that both clinicians and patients have tools and approaches for collecting, interpreting, and sharing information. Therefore, some of the potential next steps in designing and implementing collaborative HIT would be to create standard terminology for daily living activities (eg, food-related practice), developing machine-readable representations for daily living activities, and visualizations of these activities for both patients and providers.

Our analysis revealed patient-generated strategies for informing collaborative HIT design [17], but each strategy has attendant risks with the potential for unintended consequences. For this reason, these strategies should be vetted and implemented in consultation with clinicians, moreover because CTM (as we reported in the Results section) requires seamless, cross-setting information flow. HIT should be designed to be activated by patients and clinicians alike.

Owing to its outsized impact on patients' decision-making processes, the social environment can both open and close gaps in chronic condition care and its maintenance. As a result, HIT should capture and communicate relevant social environmental information to clinicians to inform therapy plans [57,58].

Developing routines was reported as an effective strategy by both clinicians and patient participants. Today's consumer technologies (eg, smartphones) partially support establishing routines; however, clinicians' understanding of and contribution to these routines is limited, and some clinicians' suggestions may not work for all individuals. Future collaborative HIT (eg, sensor-based technologies) should and can support clinicians' understanding of patients' existing routines and provide therapy plan options that can better fit in the context of patients' home environments. Moreover, as both patients and clinicians report that routines are effective, new technologies can help generate routines to support health management in daily living settings. Collaborative HIT can play a role in identifying the reasons for the formation of gaps, monitoring the consequences of gaps, and supporting the implementation of gap-mitigating strategies.

An understanding of individuals' daily living contexts can help providers better support patients' chronic disease management by eliminating the 3 reasons for gap formation, as also highlighted by previous research [4,5,17,20]. Collaborative HIT can link currently disconnected clinical and consumer technologies to flesh out contextual factors [59] and attempt to neutralize potential discrepancies in social determinants of health.

The strength of this study is its theory-driven approach. The *Infinicare* framework informed data collection or analysis and focused the researchers on the cross-setting relationships among health-related activities. This study demonstrates the utility of *Infinicare* in both examining clinician-patient collaboration and informing collaborative HIT design and implementation. Moreover, this study extended the concept of the gap by providing a case showing how gaps can inform HIT design in the context of anticoagulation therapy. Another strength of this study is its use of multiple data collection methods. We used tablet-based journals complemented by interviews. Journaling allowed for moment-by-moment collection of critical data about patients' health management activities, and interviews provided rich information about patients' contexts.

Limitations and Future Work

This study is limited in that it drew data from a single outpatient anticoagulation clinic that uses warfarin and does not have home monitoring (nonetheless, this is likely a typical US anticoagulation clinic). Although we captured rich information on the causes and consequences of gaps and strategies for avoiding them, other clinical settings may reveal additional gaps. In addition, the study design leaned toward the patient's perspective rather than the clinician's perspective (eg, discussion of gap-related consequences did not address clinician workload). It is likely that there are more consequences to clinicians (eg, in terms of workload, patient education) than we reported. In our future work, we will seek to better understand the gaps from a clinician's perspective. To minimize researcher bias, we collected data from 3 resources and triangulated them across data sources. Data were collected and analyzed by multiple researchers.

Conclusions

The study used multiple methods to explore the idea of gaps in a way to inform collaborative HIT. When designing and implementing these technologies, health-related activities conducted in daily living settings and clinical activities should be well connected and together they should be examined as a single workflow. Collaborative HIT can support clinicians and health systems to make decisions toward narrowing or eliminating the gaps can potentially improve patient and organizational outcomes.

Acknowledgments

This study was supported by the Agency for Healthcare Research and Quality under award number R03HS024092 (principal investigator: MO). The authors thank Suzanne Lareau and Jason Weiss for editorial support.

Conflicts of Interest

None declared.

References

1. Brennan PF, Casper G. Observing health in everyday living: ODLs and the care-between-the-care. Pers Ubiquit Comput 2014 Aug 11;19(1):3-8. [doi: 10.1007/s00779-014-0805-0]
2. Holden RJ, Schubert CC, Mickelson RS. The patient work system: an analysis of self-care performance barriers among elderly heart failure patients and their informal caregivers. Appl Ergon 2015 Mar;47:133-150 [FREE Full text] [doi: 10.1016/j.apergo.2014.09.009] [Medline: 25479983]
3. Nunes F, Fitzpatrick G. Understanding the mundane nature of self-care: ethnographic accounts of people living with parkinson's. In: Proceedings of the. 2018 Presented at: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; 2018; Montreal QC, Canada p. E. [doi: 10.1145/3173574.3173976]
4. Ozkaynak M, Valdez R, Holden RJ, Weiss J. Infinicare framework for integrated understanding of health-related activities in clinical and daily-living contexts. Health Syst (Basingstoke) 2018;7(1):66-78 [FREE Full text] [doi: 10.1080/20476965.2017.1390060] [Medline: 31214339]
5. Rosenfeld D, Weinberg D. Domestic practice, situated contingency and adherence to medical directives: a call for research. Soc Theory Health 2011 Jun 22;10(1):42-60. [doi: 10.1057/sth.2011.9]
6. Ozkaynak M, Jones J, Weiss J, Klem P, Reeder B. A workflow framework for health management in daily living settings. Stud Health Technol Inform 2016;225:392-396. [Medline: 27332229]
7. Ozkaynak M, Johnson S, Tulu B, Donovan J, Kanaan A, Rose A. Exploring the effect of complex patients on care delivery tasks. Int J Health Care Qual Assur 2015;28(5):494-509. [doi: 10.1108/IJHCQA-11-2014-0105] [Medline: 26020430]
8. Booth AO, Lowis C, Dean M, Hunter SJ, McKinley MC. Diet and physical activity in the self-management of type 2 diabetes: barriers and facilitators identified by patients and health professionals. Prim Health Care Res Dev 2013 Jul;14(3):293-306. [doi: 10.1017/S1463423612000412] [Medline: 23739524]
9. Arora VM, Prochaska ML, Farnan JM, D'Arcy MJ, Schwanz KJ, Vinci LM, et al. Problems after discharge and understanding of communication with their primary care physicians among hospitalized seniors: a mixed methods study. J Hosp Med 2010 Sep;5(7):385-391 [FREE Full text] [doi: 10.1002/jhm.668] [Medline: 20578045]
10. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. J Am Med Assoc 2007 Mar 28;297(8):831-841. [doi: 10.1001/jama.297.8.831] [Medline: 17327525]
11. Boissy A, Windover AK, Bokar D, Karafa M, Neuendorf K, Frankel RM, et al. Communication skills training for physicians improves patient satisfaction. J Gen Intern Med 2016 Jul;31(7):755-761 [FREE Full text] [doi: 10.1007/s11606-016-3597-2] [Medline: 26921153]
12. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Mar 24;17(2):52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
13. McInnes DK, Shimada SL, Midboe AM, Nazi KM, Zhao S, Wu J, et al. Patient use of electronic prescription refill and secure messaging and its association with undetectable hiv viral load: a retrospective cohort study. J Med Internet Res 2017 Feb 15;19(2):34 [FREE Full text] [doi: 10.2196/jmir.6932] [Medline: 28202428]
14. Schillinger D, Piette J, Grumbach K, Wang F, Wilson C, Daher C, et al. Closing the loop: physician communication with diabetic patients who have low health literacy. Arch Intern Med 2003 Jan 13;163(1):83-90. [doi: 10.1001/archinte.163.1.83] [Medline: 12523921]
15. Ozkaynak M, Brennan PF, Hanauer DA, Johnson S, Aarts J, Zheng K, et al. Patient-centered care requires a patient-oriented workflow model. J Am Med Inform Assoc 2013 Jun;20(1):e14-e16 [FREE Full text] [doi: 10.1136/amiajnl-2013-001633] [Medline: 23538724]
16. Werner NE, Malkana S, Gurses AP, Leff B, Arbaje AI. Toward a process-level view of distributed healthcare tasks: medication management as a case study. Appl Ergon 2017 Nov;65:255-268. [doi: 10.1016/j.apergo.2017.06.020] [Medline: 28802446]
17. Senteio C, Veinot T. Trying to make things right: adherence work in high-poverty, African American neighborhoods. Qual Health Res 2014 Dec;24(12):1745-1756. [doi: 10.1177/1049732314549027] [Medline: 25212857]
18. Greenwald JL, Denham CR, Jack BW. The hospital discharge: a review of a high risk care transition with highlights of a reengineered discharge process. J of Patient Safety 2007;3(2):97-106. [doi: 10.1097/01.jps.0000236916.94696.12]
19. Kripalani S, Jackson AT, Schnipper JL, Coleman EA. Promoting effective transitions of care at hospital discharge: a review of key issues for hospitalists. J Hosp Med 2007 Sep;2(5):314-323. [doi: 10.1002/jhm.228] [Medline: 17935242]
20. Kaziunas E, Ackerman MS, Veinot TC. Localizing chronic disease management: information work and health translations. In: Proc Am Soc Info Sci Tech. 2014 May 08 Presented at: Proceedings of the American Society for Information Science and Technology ;50(1) Online Firstpub Date|; 2013; Montreal, Quebec, Canada p. 1-10. [doi: 10.1002/meet.14505001090]

21. Dedhia P, Kravet S, Bulger J, Hinson T, Sridharan A, Kolodner K, et al. A quality improvement intervention to facilitate the transition of older adults from three hospitals back to their homes. J Am Geriatr Soc 2009 Sep;57(9):1540-1546. [doi: 10.1111/j.1532-5415.2009.02430.x] [Medline: 19694865]
22. Shimada S, Petrakis B, Rothendler J, Zirkle M, Zhao S, Feng H, et al. An analysis of patient-provider secure messaging at two Veterans Health Administration medical centers: message content and resolution through secure messaging. J Am Med Inform Assoc 2017 Sep 01;24(5):942-949 [FREE Full text] [doi: 10.1093/jamia/ocx021] [Medline: 28371896]
23. Chhabra PT, Rattinger GB, Dutcher SK, Hare ME, Parsons KL, Zuckerman IH. Medication reconciliation during the transition to and from long-term care settings: a systematic review. Res Social Adm Pharm 2012;8(1):60-75. [doi: 10.1016/j.sapharm.2010.12.002] [Medline: 21511543]
24. Prvu Bettger J, Alexander KP, Dolor RJ, Olson DM, Kendrick AS, Wing L, et al. Transitional care after hospitalization for acute stroke or myocardial infarction: a systematic review. Ann Intern Med 2012 Sep 18;157(6):407-416 [FREE Full text] [doi: 10.7326/0003-4819-157-6-201209180-00004] [Medline: 22986378]
25. Manderson B, McMurray J, Piraino E, Stolee P. Navigation roles support chronically ill older adults through healthcare transitions: a systematic review of the literature. Health Soc Care Community 2012 Mar;20(2):113-127. [doi: 10.1111/j.1365-2524.2011.01032.x] [Medline: 21995806]
26. Albert NM. A systematic review of transitional-care strategies to reduce rehospitalization in patients with heart failure. Heart Lung 2016;45(2):100-113. [doi: 10.1016/j.hrtlng.2015.12.001] [Medline: 26831374]
27. Hirschman KB, Shaid E, McCauley K, Pauly MV, Naylor MD. Continuity of care: the transitional care model. Online J Issues Nurs 2015 Sep 30;20(3):1 [FREE Full text] [Medline: 26882510]
28. Miller AR, Condin CJ, McKellin WH, Shaw N, Klassen AF, Sheps S. Continuity of care for children with complex chronic health conditions: parents' perspectives. BMC Health Serv Res 2009 Dec 21;9:242 [FREE Full text] [doi: 10.1186/1472-6963-9-242] [Medline: 20025770]
29. Klasnja P, Hartzler A, Unruh K, Pratt W. Blowing in the wind: unanchored patient information work during cancer care. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Blowing in the Wind: Unanchored Patient Information Work during Cancer Care. Proc SIGCHI Conf Hum Factor Comput Syst 2010; 2010 Apr Presented at: CHI 2010; April 10-15, 2015; Atlanta, GA p. 193-202 URL: http://europepmc.org/abstract/MED/20697453 [doi: 10.1145/1753326.1753355]
30. Dixon B, Embi P, Haggstrom D. Information technologies that facilitate care coordination: provider and patient perspectives. Transl Behav Med 2018 May 23;8(3):522-525. [doi: 10.1093/tbm/ibx086] [Medline: 29800411]
31. Feldman S, Horan T. The dynamics of information collaboration: a case study of blended IT value propositions for health information exchange in disability determination. J Asso for Info Sys 2011 Feb;12(2):189-207. [doi: 10.17705/1jais.00256]
32. Valdez R, Holden R, Novak L, Veinot T. Technical infrastructure implications of the patient work framework. J Am Med Inform Assoc 2015 Apr;22(e1):213-215 [FREE Full text] [doi: 10.1093/jamia/ocu031] [Medline: 25665705]
33. Samaras GM, Horst RL. A systems engineering perspective on the human-centered design of health information systems. J Biomed Inform 2005 Mar;38(1):61-74 [FREE Full text] [doi: 10.1016/j.jbi.2004.11.013] [Medline: 15694886]
34. Strauss AT, Martinez DA, Garcia-Arce A, Taylor S, Mateja C, Fabri PJ, et al. A user needs assessment to inform health information exchange design and implementation. BMC Med Inform Decis Mak 2015 Oct 12;15:81 [FREE Full text] [doi: 10.1186/s12911-015-0207-x] [Medline: 26459258]
35. Richardson J, Vest J, Green C, Kern L, Kaushal R, HITEC Investigators. A needs assessment of health information technology for improving care coordination in three leading patient-centered medical homes. J Am Med Inform Assoc 2015 Jul;22(4):815-820. [doi: 10.1093/jamia/ocu039] [Medline: 25796597]
36. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
37. Chung C, Dew K, Cole A, Zia J, Fogarty J, Kientz JA, et al. Boundary negotiating artifacts in personal informatics: patient-provider collaboration with patient-generated data. In: CSCW Conf Comput Support Coop Work. 2016 Feb 27 Presented at: Comput Support Coop Work; February 27–March 2, 2016; San Francisco, CA p. 770-786 URL: http://europepmc.org/abstract/MED/28516171 [doi: 10.1145/2818048.2819926]
38. O'Brien R, Wyke S, Watt GG, Guthrie B, Mercer SW. The 'everyday work' of living with multimorbidity in socioeconomically deprived areas of Scotland. J Comorb 2014;4:1-10 [FREE Full text] [doi: 10.15256/joc.2014.4.32] [Medline: 29090148]
39. Schroeder J, Hoffswell J, Chung C, Fogarty J, Munson S, Zia J. Supporting patient-provider collaboration to identify individual triggers using food and symptom journals. In: CSCW Conf Comput Support Coop Work. 2017 Feb 25 Presented at: CSCW Conf Comput Support Coop Work; February 25–March 1, 2017; Portland, OR p. 1726-1739 URL: http://europepmc.org/abstract/MED/28516172 [doi: 10.1145/2998181.2998276]
40. Ansell J, Hirsh J, Hylek E, Jacobson A, Crowther M, Palareti G. Pharmacology and management of the Vitamin K antagonists: American College of Chest Physicians evidence-based clinical practice guidelines (8th edition). Chest 2008 Jun;133(6 Suppl):160-198. [doi: 10.1378/chest.08-0670] [Medline: 18574265]
41. Ansell JE, Oertel LB, Wittkowsky AK, editors. Managing oral anticoagulation therapy: clinical and operational guidelines. St. Louis, MO: Wolters Kluwer Health; 2009.

42. Boyatzis RE. Transforming qualitative information: thematic analysisand code development. Thousand Oaks, CA: Sage Publications; 1998:-.
43. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
44. Graneheim U, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Mar;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
45. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
46. Barbour R. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? Br Med J 2001 May 05;322(7294):1115-1117 [FREE Full text] [doi: 10.1136/bmj.322.7294.1115] [Medline: 11337448]
47. Saldana J. The coding manual for qualitative researchers. Los Angeles: SAGE Publications; 2013:-.
48. Kogan AC, Wilber K, Mosqueda L. Person-centered care for older adults with chronic conditions and functional impairment: a systematic literature review. J Am Geriatr Soc 2016 Jan;64(1):1-7. [doi: 10.1111/jgs.13873] [Medline: 26626408]
49. Granger BB, Sandelowski M, Tahshjain H, Swedberg K, Ekman I. A qualitative descriptive study of the work of adherence to a chronic heart failure regimen: patient and physician perspectives. J Cardiovasc Nurs 2009;24(4):308-315. [doi: 10.1097/JCN.0b013e3181a4be30] [Medline: 19465865]
50. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 2015 Apr 09;13:77 [FREE Full text] [doi: 10.1186/s12916-015-0319-2] [Medline: 25889598]
51. Chung C, Cook J, Bales E, Zia J, Munson SA. More than telemonitoring: health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. J Med Internet Res 2015 Aug 21;17(8):203 [FREE Full text] [doi: 10.2196/jmir.4364] [Medline: 26297627]
52. Hull S. Patient-generated health data foundation for personalized collaborative care. Comput Inform Nurs 2015 May;33(5):177-180. [doi: 10.1097/CIN.0000000000000159] [Medline: 25974360]
53. Woods S, Evans N, Frisbee K. Integrating patient voices into health information for self-care and patient-clinician partnerships: Veterans Affairs design recommendations for patient-generated data applications. J Am Med Inform Assoc 2016 May;23(3):491-495. [doi: 10.1093/jamia/ocv199] [Medline: 26911810]
54. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. Addict Behav 2018 Aug;83:42-47 [FREE Full text] [doi: 10.1016/j.addbeh.2017.11.039] [Medline: 29217132]
55. Boushey CJ, Spoden M, Zhu FM, Delp EJ, Kerr DA. New mobile methods for dietary assessment: review of image-assisted and image-based dietary assessment methods. In: Proc Nutr Soc. 2017 Presented at: Nutrition Society Summer Meeting; July 11–14, 2016; Dublin, Ireland p. 283-294. [doi: 10.1017/s0029665116002913]
56. Lobelo F, Kelli HM, Tejedor SC, Pratt M, McConnell MV, Martin SS, et al. The wild wild west: a framework to integrate mhealth software applications and wearables to support physical activity assessment, counseling and interventions for cardiovascular disease risk reduction. Prog Cardiovasc Dis 2016;58(6):584-594 [FREE Full text] [doi: 10.1016/j.pcad.2016.02.007] [Medline: 26923067]
57. Menefee HK, Thompson MJ, Guterbock TM, Williams IC, Valdez RS. Mechanisms of communicating health information through Facebook: implications for consumer health information technology design. J Med Internet Res 2016 Aug 11;18(8):218 [FREE Full text] [doi: 10.2196/jmir.5949] [Medline: 27515151]
58. Valdez RS, Brennan PF. Exploring patients' health information communication practices with social network members as a foundation for consumer health IT design. Int J Med Inform 2015 May;84(5):363-374. [doi: 10.1016/j.ijmedinf.2015.01.014] [Medline: 25704761]
59. Unruh K, Pratt W. The invisible work of being a patient and implications for health care: “[the doctor is] my business partner in the most important business in my life, staying alive.”. In: Conf Proc Ethnogr Prax Ind Conf. 2008 Nov Presented at: Ethnogr Prax Ind Conf 2008; October 15 -18, 2008; Copenhagen, Denmark p. 40-50 URL: http://europepmc.org/abstract/MED/21709742 [doi: 10.1111/j.1559-8918.2008.tb00093.x]

Abbreviations

CTM: collaborative therapy management
EHR: electronic health record
HIT: health information technology
INR: international normalized ratio
IT: information technology

Edited by G Eysenbach, Q Zeng; submitted 03.01.20; peer-reviewed by A Lau, MH Chung, K Day, L Vizer, D Wu; comments to author 28.08.20; revised version received 24.11.20; accepted 18.01.21; published 25.02.21.

Please cite as:
Ozkaynak M, Valdez R, Hannah K, Woodhouse G, Klem P
Understanding Gaps Between Daily Living and Clinical Settings in Chronic Disease Management: Qualitative Study
J Med Internet Res 2021;23(2):e17590
URL: https://www.jmir.org/2021/2/e17590
doi:10.2196/17590
PMID:33629657

Original Paper

Differences in the Gut Microbiome of Women With and Without Hypoactive Sexual Desire Disorder: Case Control Study

Guanjian Li[1*], MD; Weiran Li[2*], MD, PhD; Bing Song[3*], MD, PhD; Chao Wang[4*], MD; Qunshan Shen[5*], MD; Bo Li[6], MD, PhD; Dongdong Tang[7], MD; Chuan Xu[2], MD; Hao Geng[7]; Yang Gao[3], MD; Guanxiong Wang[3], MD; Huan Wu[3], MD; Zhiguo Zhang[1], MD, PhD; Xiaofeng Xu[7], MD, PhD; Ping Zhou[3], MD, PhD; Zhaolian Wei[5], MD, PhD; Xiaojin He[4*], MD, PhD; Yunxia Cao[1*], MD, PhD

[1]Reproductive Medicine Center, Department of Obstetrics and Gynecology, The First Affiliated Hospital of Anhui Medical University, Hefei, China
[2]The First Affiliated Hospital of Anhui Medical University, Hefei, China
[3]National Health Commission Key Laboratory of Study on Abnormal Gametes and Reproductive Tract, Hefei, China
[4]Key Laboratory of Population Health Across Life Cycle, Ministry of Education of the People's Republic of China, Hefei, China
[5]Anhui Province Key Laboratory of Reproductive Health and Genetics, Hefei, China
[6]Taikang Tongji Hospital, Wuhan, China
[7]Biopreservation and Artificial Organs, Anhui Provincial Engineering Research Center, Hefei, China
[*]these authors contributed equally

Corresponding Author:
Yunxia Cao, MD, PhD
Reproductive Medicine Center
Department of Obstetrics and Gynecology
The First Affiliated Hospital of Anhui Medical University
81 Meishan Road
Hefei,
China
Phone: 86 15395104659
Email: caoyunxia6@126.com

Abstract

Background: The gut microbiome is receiving considerable attention as a potentially modifiable risk factor and therapeutic target for numerous mental and neurological diseases.

Objective: This study aimed to explore and assess the difference in the composition of gut microbes and fecal metabolites between women with hypoactive sexual desire disorder (HSDD) and healthy controls.

Methods: We employed an online recruitment method to enroll "hard-to-reach" HSDD populations. After a stringent diagnostic and exclusion process based on DSM-IV criteria, fecal samples collected from 24 women with HSDD and 22 age-matched, healthy controls underwent microbiome analysis using 16S ribosomal RNA gene sequencing and metabolome analysis using untargeted liquid chromatography–mass spectrometry.

Results: We found a decreased abundance of *Ruminococcaceae* and increased abundance of *Bifidobacterium* and *Lactobacillus* among women with HSDD. Fecal samples from women with HSDD showed significantly altered metabolic signatures compared with healthy controls. The abundance of *Bifidobacterium*, *Lactobacillus*, and several fecal metabolites correlated negatively with the sexual desire score, while the number of *Ruminococcaceae* correlated positively with the sexual desire score in all subjects.

Conclusions: Our analysis of fecal samples from women with HSDD and healthy controls identified significantly different gut microbes and metabolic signatures. These preliminary findings could be useful for developing strategies to adjust the level of human sexual desire by modifying gut microbiota.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1800020321; http://www.chictr.org.cn/showproj.aspx?proj=34267

(J Med Internet Res 2021;23(2):e25342) doi:10.2196/25342

KEYWORDS

gut microbiome; metabolome; sexual desire; online recruitment; biomarkers

Introduction

Hypoactive sexual desire disorder (HSDD) is defined as a deficiency or absence of desire for sexual activity that causes marked distress or interpersonal difficulty and is not accounted for by another psychiatric condition, use of medications, or relationship problems [1,2]. Two large epidemiological surveys have shown that this combination of symptoms, "low desire and associated distress," is present in up to 10% of American women, with similar prevalence rates seen throughout the world [3,4]. HSDD is the most frequently reported female sexual health problem, and women with HSDD may experience reduced quality of life, impaired physical image, and decreased feelings of self-confidence and self-worth, in addition to feeling less connected with their partners and families [5].

In humans, sexual desire is regulated by key areas of the brain through the action of various neurotransmitters [6,7]. Norepinephrine, dopamine, melanocortin, oxytocin, and vasopressin mediate sexual excitation, while serotonin, opioids, prolactin, and the endogenous cannabinoid system mediate sexual inhibition. Generalized HSDD may be related to a neuropsychological state of increased inhibition or decreased excitation, or a mixture of the two [8,9]. Neurotransmitters in the central nervous system are thus therapeutic targets for improving human sexual desire. At present, the main treatment strategies are to reduce the action of 5-HT, enhance the action of dopamine, or both [10].

Human beings exist in a mutualistic relationship with their gut microbiota, a highly diverse and complex microbial ecosystem closely associated with various phenotypes and diseases [11,12]. Perturbations in gut microbe richness and diversity affect the levels of 5-HT, norepinephrine, and γ-aminobutyric acid (GABA)ergic and dopaminergic neurotransmission in the brain [13-15]. These pathways and molecules are believed to be closely related to human sexual desire.

Although this evidence suggests that there may be a connection between the composition of gut microbiota and sexual desire, our knowledge of the role of gut microbiota in sexual desire disorders is limited. We hypothesized that the fecal gut microbiota and metabolites of patients with sexual desire disorders differ from those of people without such disorders and that active metabolites and neurochemicals may be mediators between gut microbes and the sexual desire system.

Recruiting participants for microbial research is increasingly difficult as the number of projects competing for participants' attention increases and response rates decline. In addition, women with desire disorder constitute a hard-to-reach population owing to fear and lack of trust in the study procedures or in research staff, especially in a more conservative society like China when compared with Western countries [16]. Widespread access to online tools now offers various advantages in data collection that can reduce administrative procedures and improve information privacy [17,18]. Previous studies have demonstrated that performance and effectiveness obtained using online recruitment methods in studies of sexual issues or gut microbes concur with those obtained using offline recruitment methods, such as telephone surveys, mail-in questionnaires, posters, and flyers [12,19-21]. In addition, online surveys may even provide greater reach and be less expensive, while allowing for convenient data collection and analysis [22-24]. These findings suggest that the growth of the internet and social media sites may provide new opportunities for research on the correlation between sensitive health concerns (such as mental health problems or sexual health disorders) and gut microbes.

We conducted this study to investigate the composition of the gut microbiome and metabolite abundance in women with HSDD and to increase our understanding of the possible association between human sexual desire and gut microbiota and fecal metabolites.

Methods

Participants

From March 2019 to November 2019, we recruited subjects by posting advertisements on instant messenger applications and the information platform. In all advertisements, potential participants were automatically redirected to a study website that included a link to the survey and participant information including study details and information statement.

Potential recruits included women who had previously been diagnosed with suspected HSDD but not yet treated or who complained of decreased sexual desire, causing them to seek medical counseling, as well as volunteers who claimed to have normal sexual desire. No incentives were offered to responders. When requested, they were provided a brief report on their own gut microbial composition. All responders provided electronic written informed consent. All experimental protocols used in this study were approved by the Ethics Committee of Anhui Medical University and the relevant hospitals.

Clinical Information Collection

We collected information online from 157 premenopausal women and scheduled face-to-face diagnostic interviews and biological sample collection to identify women with generalized acquired HSDD. Incomplete questionnaires and responses that rejected further face-to-face interviews were not accepted by the system. A combination of a web-based questionnaire and structured interviews was used to collect information on demographic characteristics such as ethnicity, marital status, income, educational level, reproductive history, sexual partner relationships, sexual frequency, sexual dysfunction, and the presence and treatment of related diseases.

16S Ribosomal RNA Gene Sequencing and Metabolomics Tests

Subsequently, we performed high-throughput 16S ribosomal RNA gene sequencing and untargeted liquid chromatography–mass spectrometry metabolomic analysis of stool samples from the 26-member HSDD cohort and 26 healthy controls (for all survey measures and analysis methods, see the Methods section in Multimedia Appendix 1).

Results

Clinical Characteristics of the Subjects

A flowchart showing the inclusion and exclusion processes is presented in Figure 1. In the end, 24 women with HSDD and 22 women with no history of sexual dysfunction (NHSD) were included in the final analyses. The clinicopathological variables of the HSDD and NHSD cohorts are shown in Table 1. All participants were Han Chinese from the Hefei area and had comparable geographical conditions and eating habits. The clinicopathological variables of the HSDD and NHSD cohorts were comparable, except for a slight increase in platelet count in the HSDD cohort (the meaning is unclear). Significant differences were observed in Female Sexual Function Index desire domain scores between the HSDD and NHSD groups.

Figure 1. Inclusion and exclusion processes. HSDD: hypoactive sexual desire disorder; LC-MS: liquid chromatography–mass spectrometry; NHSD: no history of sexual dysfunction.

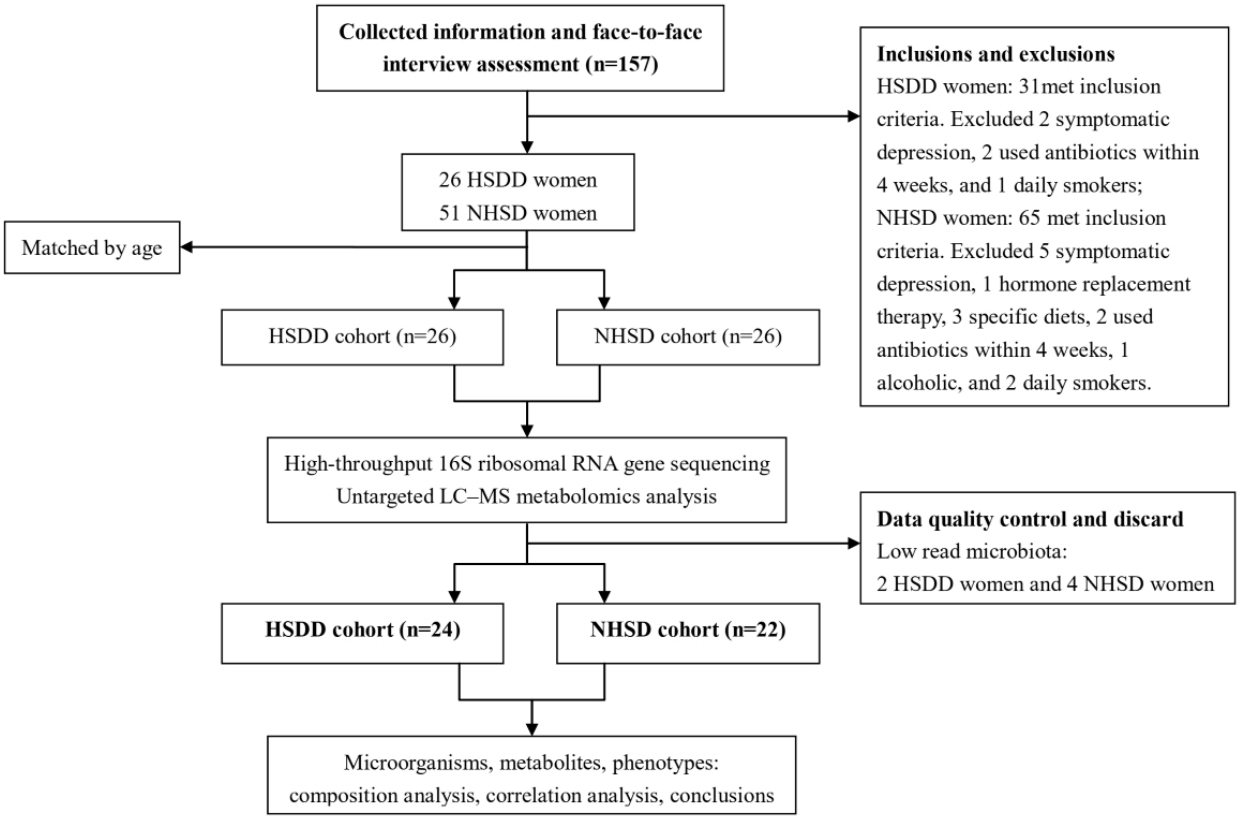


Table 1. Demographic and other sample data at baseline.

Characteristic	HSDD[a] (n=24)	NHSD[b] (n=22)	*P* value
Age (years), mean (SD)	33.8 (4.5)	33.9 (4.0)	.91
BMI (kg/m^2), mean (SD)	27.1 (3.6)	26.3 (3.3)	.27
College or beyond, n	17	16	.89
FSFI-D[c] score, mean (SD)	3.2 (0.9)	8.6 (1.0)	.01
CES-D[d] score, mean (SD)	10.3 (4.9)	10.9 (4.6)	.55
RBC[e] (×10^{12}/L), mean (SD)	4.3 (0.7)	4.1 (0.4)	.06
WBC[f] (×10^9/L), mean (SD)	6.9 (1.7)	6.5 (1.7)	.28
PLT[g] (×10^{12}/L), mean (SD)	169.4 (44.8)	154.0 (35.9)	.04
FSH[h,i] (mIU/mL), mean (SD)	6.3 (3.0)	6.3 (2.9)	.93
LH[i,j] (mIU/mL), mean (SD)	5.3 (3.4)	5.9 (3.7)	.67
Estradiol[i] (pg/mL), mean (SD)	48.1(9.3)	50.6 (6.5)	.09
Total T[i,k] (ng/mL), mean (SD)	0.5 (0.3)	0.5 (0.2)	.86

[a]HSDD: hypoactive sexual desire disorder.

[b]NHSD: no history of sexual dysfunction.

[c]FSFI-D: Female Sexual Function Index desire domain.

[d]CES-D: Center for Epidemiologic Studies Depression Scale.

[e]RBC: red blood cell.

[f]WBC: white blood cell.

[g]PLT: platelet.

[h]FSH: follicle-stimulating hormone.

[i]Due to missing data, the group sizes differ for hormonal data: HSDD group, n=20; NHSD group, n=14.

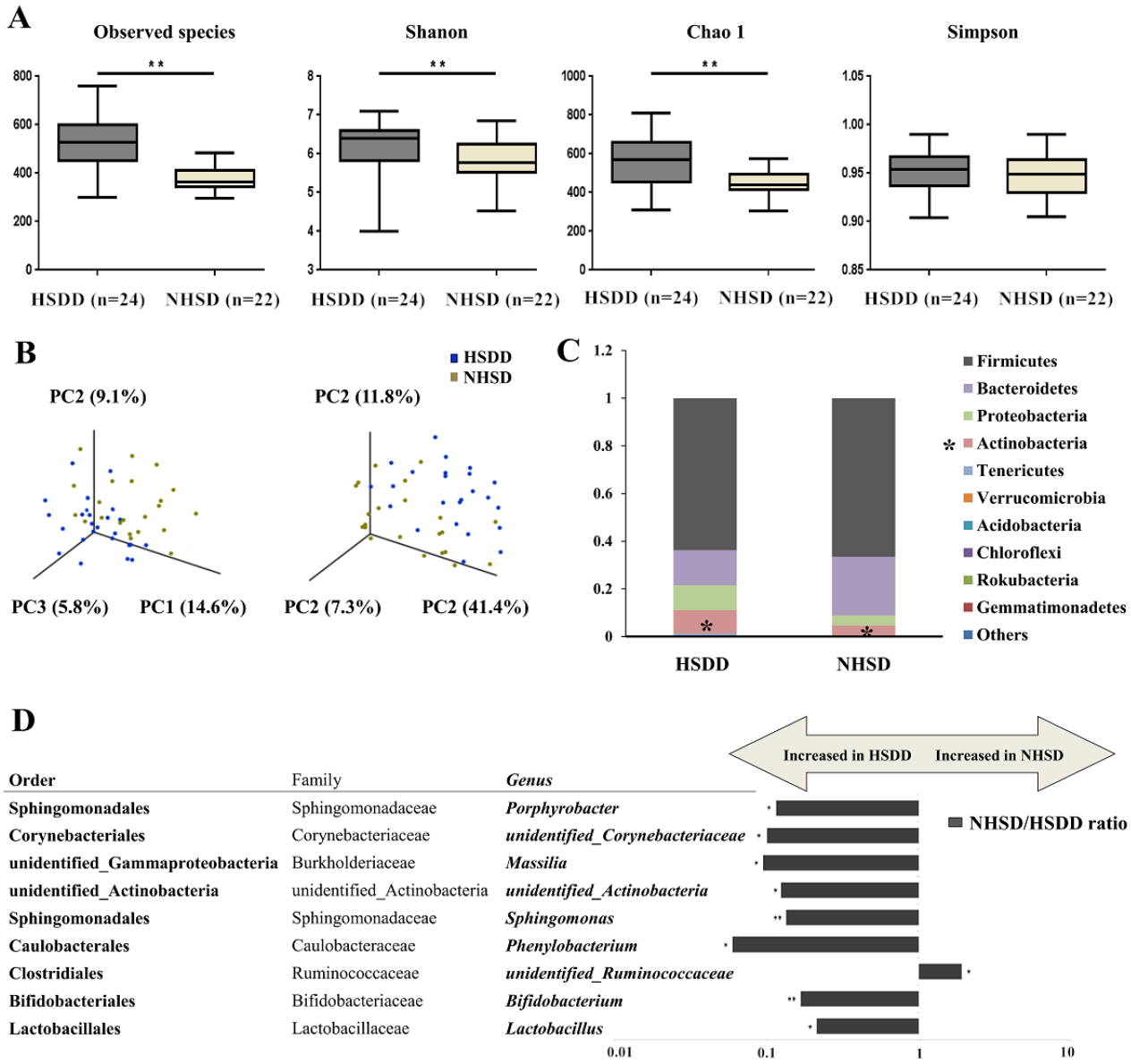
[j]LH: luteinizing hormone.

[k]T: testosterone.

HSDD and Gut Microorganisms

In our investigation, 1753 operational taxonomic units were identified from sequenced specimens with 97% sequence similarity. Different operational taxonomic unit–based diversity indexes were used to assess the bacterial composition within each sample (the α-diversity). Compared to the NHSD group, the HSDD cohort was characterized by a higher number of observed species and higher Shannon and Chao 1 indexes (Figure 2A, except for Simpson), suggesting greater species diversity in gut microbiota in the HSDD cohort. β-diversity analysis was performed to determine whether HSDD was associated with altered gut microbial composition. As shown in Figure 2B, both unweighted and weighted principal coordinate analysis plots revealed that the gut microbial composition of the HSDD group was significantly different from that of the NHSD group.

Figure 2. Gut microbial characteristics of women with hypoactive sexual desire disorder (HSDD) or with no history of sexual dysfunction (NHSD), based on (A) α-diversity analysis of 3 indexes (observed species, shannon, and Chao 1); (B) β-diversity analysis, providing unweighted and weighted principal coordinate (PC) plots; (C) component proportion of bacterial phylum in each group; (D) the ratio between HSDD and NHSD on the logarithmic scale. All QFDR<0.05.



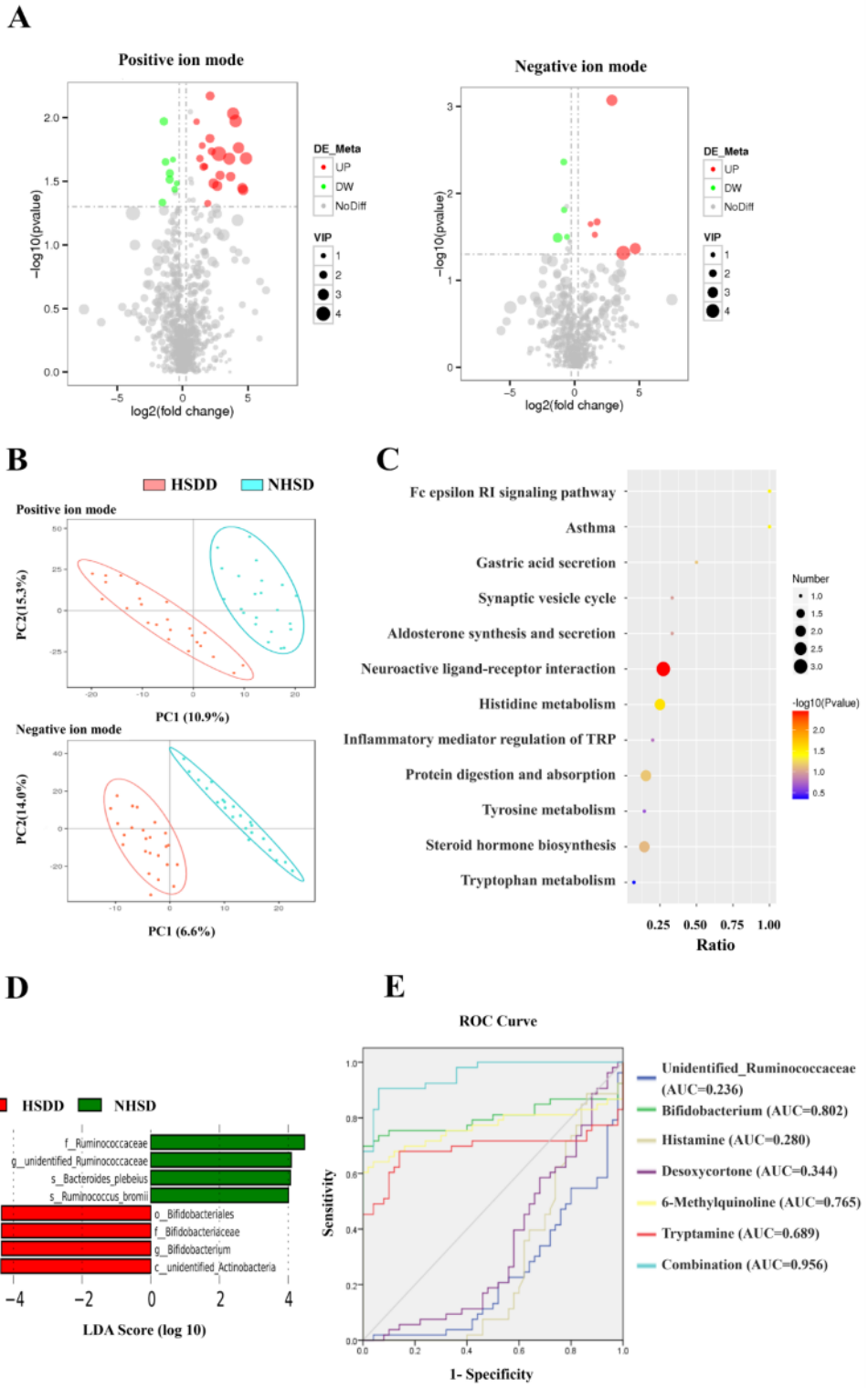
At the phylum level, the most abundant taxa were *Firmicutes* and *Bacteroidetes*, followed by *Proteobacteria* and *Actinobacteria* in the exploration cohort (Figure 2C). *Actinobacteria* were significantly enriched in the HSDD group compared to the NHSD group (9.1% versus 3.3%, *P*<.01). At the genus level, 9 taxa were different between the HSDD and NHSD groups (Figure 2D). *Bifidobacterium*, *Lactobacillus*, and *Sphingomonas* were enriched in women with HSDD, while an unidentified genus of *Ruminococcaceae* was enriched in NHSD women (*P*<.05 and QFDR<0.20). *Lactobacillus* and *Ruminococcaceae* belong to the same phylum, *Firmicutes*, which is the most abundant phylum, while *Bifidobacterium* belong to the phylum *Actinobacteria*.

HSDD and the Fecal Metabolome

In the fecal metabolome analysis, 1168 metabolites were quantified. Differential metabolite volcanic maps show that the fecal metabolic phenotype of women with HSDD was significantly different than that of women with NHSD (Figure 3A). Principal component analysis clearly showed differences between the HSDD and NHSD groups, indicating that women with HSDD have specific metabolic abundances and characteristics (Figure 3B). In total, 28 metabolites that made major contributions to the distinction (VIP score >1.5) were selected using partial least squares discriminant analysis. These differential metabolites were annotated to 12 different KEGG metabolic pathways [25], including neuroactive ligand-receptor interactions, the Fc epsilon RI signaling pathway, steroid

hormone biosynthesis, histidine metabolism, and tryptophan metabolism (Figure 3C).

Figure 3. Fecal metabolites characteristics of women with hypoactive sexual desire disorder (HSDD) and women with no history of sexual dysfunction (NHSD), displayed as (A) volcano plots showing the differentially accumulated (log2[fold change] and significantly changed (-log10[pvalue] fecal metabolites; (B) principal component (PC) analysis showing grouped discrimination; (C) functions and pathways of these metabolites using the KEGG database, where the differential metabolites are annotated on 12 different KEGG metabolic pathways; (D) Lefse and linear discriminant analyses showing differences in taxonomic composition; (E) receiver operating characteristic (ROC) curve. AUC: area under the curve; TRP: tryptophan.



A Panel of Fecal Microorganisms and Metabolic Markers Identifies HSDD

To further identify the specific bacterial components associated with HSDD, we used a standard microbiota analysis tool, LefSe; 8 taxa were identified as key biomarkers (P<.05, linear discriminant analysis scores >4). Several taxa, including the genus *Bifidobacterium* and class unidentified *Actinobacteria*, were significantly overrepresented in the HSDD women, whereas the genus unidentified *Ruminococcaceae* was overrepresented in the NHSD women (Figure 3D).

To identify and quantify the potential of microbial and metabolic biomarkers for HSDD, we performed a receiver operating characteristic curve analysis using the relative abundance of the *Bifidobacterium* genus and the unidentified_*Ruminococcaceae* genus, the most abundant genus identified, to distinguish HSDD from NHSD. Then, a binary logistic regression analysis for the 2 genera and 28 significantly correlated metabolites was performed. Significant deviations between women with HSDD and women with NHSD could be identified using *Bifidobacterium* and unidentified_*Ruminococcaceae* at the genus level, as well as 4 significantly correlated metabolites: histamine (P=.01), desoxycortone (P=.03), 6-Methylquinoline (P=.01), and tryptamine (P=.03). This diagnostic marker panel (combination of 2 microbial markers and 4 fecal metabolites) discriminated between the HSDD and NHSD groups with 95.6% accuracy (area under the curve=0.956; Figure 3E).

Relationships Between Gut Microbiota, Metabolites, and Clinical Characteristics

A Spearman's rank correlation test was performed to explore the correlations between gut microbes, fecal metabolites, and clinical characteristics. The altered bacterial genera were generally associated with differential metabolites (Multimedia Appendix 2), and almost all the differential bacterial genera showed significant correlations with a range of differential fecal metabolites (P<.05). Spearman correlation analysis between every 2 variables suggested that some bacterial genera and fecal metabolomes were slightly associated with clinical characteristics such as age, BMI, Center for Epidemiologic Studies Depression Scale score, and some blood indicators (Multimedia Appendix 2). We found that some genera and metabolites were significantly correlated with the Female Sexual Function Index desire domain score ($|r|>0.35$, P<.05), providing strong evidence for the existence of a correlation between sexual desire and gut microorganisms.

Discussion

Principal Findings

These results suggest significant differences in the composition of fecal microbiota between controls and women with HSDD. We observed a significantly higher quantity of *Actinobacteria* in women with HSDD compared with healthy controls. The altered abundance of several genera contributed to the unique gut microbial characteristics found in HSDD. The most prominent of these was the marked enrichment of the *Bifidobacterium* genus. Correlation analysis showed that the greater number of *Bifidobacterium* and *Lactobacillus* could be associated with a decrease in sexual desire. Higher microbial diversity has been shown in previous studies to have benefits for the human body, and elevated levels of *Bifidobacterium* and *Lactobacillus* have been related to reduced anger, dysphoric mood, aggressive thoughts, and self-reported feelings of sadness [26,27]. A possible explanation for these observations is that these emotions and human sexual desires are physiologically interrelated, involving a variety of neurotransmitters and other changes in brain biochemistry. Anger, fear, or stress can be a prelude to sex, specifically because these emotional states generate enthusiasm or excitement that leads to an "excitement transfer" into sexual desire and arousal [28].

There is considerable evidence that gut microorganisms modulate the metabolism of dopamine, 5-HT, and noradrenaline in the brain [15,29,30]. These molecules are generally considered to be the major neurotransmitters regulating sexual desire [7]. Previous research suggests that plasma concentrations of kynurenic acid and tryptophan (serotonergic precursor) in Sprague-Dawley rats treated with *Bifidobacterium* for 14 days increased significantly compared with controls. *Bifidobacterium* treatment also resulted in a decrease in 5-hydroxyindoleacetic acid concentration in the frontal cortex and dihydroxyphenylacetic acid concentration in the amygdala [31]. In another study, rats were orally administered specific *Bifidobacterium* every day, and *Bifidobacterium* reportedly produced GABA through glutamate decarboxylase [32]. *Ruminococcaceae*, also commonly referred to as Clostridia clusters XIVa and IV, are abundant microorganisms in the normal gut. They decompose indigestible carbohydrates and produce short-chain fatty acids (SCFAs). Numerous studies have shown that SCFAs, including butyrate, can cross the blood-brain barrier and enter the central nervous system and may play a role in neuropsychiatric conditions and general psychological functions [33]. SCFAs might directly influence the brain by influencing the blood-brain barrier, stimulating the vagus nerve, regulating the secretion of neurotrophic factors, modulating 5-HT/dopamine/noradrenaline/GABA biosynthesis, and promoting the transcription of inhibitory pathway transcripts [34]. These findings lead to the hypothesis that neurotransmitter molecules and pathways may play crucial roles in the regulation of sexual desire by gut microorganisms.

The analysis of fecal metabolic components has received considerable attention since these biomolecules reflect genetic and environmental effects and also act to connect the health of the host and its symbiotic microorganisms [35]. In fecal samples from women with HSDD, metabolites had significantly different distributions from the control group. Enriched functional modules included those involved in neuroactive ligand-receptor interactions, histidine metabolism, steroid hormone biosynthesis, and tryptophan metabolism.

The systematic exclusion of alternative diagnoses by structured interviews and the use of rigorous differential diagnosis criteria based on the DSM-IV resulted in a representative cohort of individuals with generalized acquired HSDD, avoiding various genital disorders existing in men or elderly women, rather than the sexual desire problem of "central effect." In addition, although we reported Center for Epidemiologic Studies Depression Scale 20 scores, estradiol, testosterone, and sex

hormone-binding globulin, in fact, the cause of HSDD is multifactorial and variable; changes in hormones, stress, depression, present sexual experiences, past history of abuse, total well-being, sexual condition of the sexual partner, and relationship factors may also be crucial causes of sexual desire disorder [36,37]. Therefore, whether our findings might be specific to HSDD or more general changes seen in multiple disorders need to be further explored in future studies with large samples.

Given the increased popularity of online media and information platforms, many researchers are using web-based tools to recruit participants for medical, psychological, and sociological research studies (especially for stigmatized and hard-to-reach groups). Previous studies have shown that online methods can successfully recruit and manage people who are humiliated or discriminated against because of race, sexual orientation, or mental health status [17,18]. Our results confirmed recent literature regarding online recruitment of hard-to-reach people and expanded the scope of recruitment and management to the study population of sexual dysfunction and gut microbe research.

Limitations

There are several potential limitations of our research. First, our online recruitment methods relied on convenience samples, and determining the source population limits the external validity of findings. When it comes to sensitive issues, determining adequate compensation and ensuring confidentiality continue to be important concerns. Second, gut microorganisms and metabolites are affected by a variety of factors, including geography, daily diet, and genetic variation [38]. Additionally, race and ethnic culture appear to have a significant association with the occurrence of sexual problems [39]. Therefore, since all the participants were from a single ethnicity, identifying specific correlations between gut microorganisms and sexual desire in a global population may not be straightforward. Another issue is that there is no standardized method for adjusting the composition of the microbiome using diet. In this study, this issue was handled by excluding all individuals who reported variations in their diet, such as binge eating or purging, following a vegetarian diet, or having gluten or lactose intolerance. However, it cannot be ruled out that subtle, unspecified dietary factors could have influenced our results. Finally, our findings only demonstrated an association between the gut microbiome and HSDD. Whether alternations in the gut microbiome are responsible for HSDD should be further validated using animal models in the laboratory.

Conclusions

This study addressed a number of gaps. We identified numerous microbial genera, including *Bifidobacterium* and *Lactobacillus*, that were enriched in women with HSDD. In addition, our metabolomic analysis of fecal samples from women with HSDD and healthy controls identified significantly different metabolic signatures. Our results support the hypothesis that disturbances in gut microorganisms are closely related to the level of human sexual desire. Pathogenicity of the microbiome in women with HSDD may be caused by comprehensive changes in holistic interaction, rather than by specific pathogens or metabolites.

A better understanding of the microbiota's role in the gut-brain-libido axis may lead to the identification of novel drug targets and management measures to improve sexual health. Much work remains to be performed in determining the role and mechanisms of the gut microbiota in human sexual desire.

Acknowledgments

The authors would like to thank all participants. We also thank the researchers from Anhui Medical University, as well as engineers, nurses, and doctors. This work was supported by the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (grant number 2019PT310002).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary methods.
[DOCX File , 28 KB - jmir_v23i2e25342_app1.docx]

Multimedia Appendix 2
Heat map. (a) Heat map of the Spearman's rank correlation coefficient of differential bacterial genus, differential metabolites and HSDD clinical characteristics. (b) Heatmaps showing altered bacterial genus were generally associated with differential fecal metabolites. The statistical significance was denoted on the squares (#p < 0.05; ##< 0.01).
[PNG File , 2362 KB - jmir_v23i2e25342_app2.png]

References

1. Diagnostic and statistical manual of mental disorders: DSM-IV. Washington DC: American Psychiatric Association; 1994.
2. Meyers M, Margraf J, Velten J. Psychological Treatment of Low Sexual Desire in Women: Protocol for a Randomized, Waitlist-Controlled Trial of Internet-Based Cognitive Behavioral and Mindfulness-Based Treatments. JMIR Res Protoc 2020 Sep 29;9(9):e20326 [FREE Full text] [doi: 10.2196/20326] [Medline: 32990248]

3. Shifren JL, Monz BU, Russo PA, Segreti A, Johannes CB. Sexual problems and distress in United States women: prevalence and correlates. Obstet Gynecol 2008 Nov;112(5):970-978. [doi: 10.1097/AOG.0b013e3181898cdb] [Medline: 18978095]
4. West SL, D'Aloisio AA, Agans RP, Kalsbeek WD, Borisov NN, Thorp JM. Prevalence of low sexual desire and hypoactive sexual desire disorder in a nationally representative sample of US women. Arch Intern Med 2008 Jul 14;168(13):1441-1449. [doi: 10.1001/archinte.168.13.1441] [Medline: 18625925]
5. Kingsberg SA. Attitudinal survey of women living with low sexual desire. J Womens Health (Larchmt) 2014 Oct;23(10):817-823. [doi: 10.1089/jwh.2014.4743] [Medline: 25105910]
6. Georgiadis JR, Kringelbach ML, Pfaus JG. Sex for fun: a synthesis of human and animal neurobiology. Nat Rev Urol 2012 Sep;9(9):486-498. [doi: 10.1038/nrurol.2012.151] [Medline: 22926422]
7. Pfaus JG. Pathways of sexual desire. J Sex Med 2009 Jun;6(6):1506-1533. [doi: 10.1111/j.1743-6109.2009.01309.x] [Medline: 19453889]
8. Toates F. An integrative theoretical framework for understanding sexual motivation, arousal, and behavior. J Sex Res 2009;46(2-3):168-193. [doi: 10.1080/00224490902747768] [Medline: 19308842]
9. Goldstein I, Kim NN, Clayton AH, DeRogatis LR, Giraldi A, Parish SJ, et al. Hypoactive Sexual Desire Disorder: International Society for the Study of Women's Sexual Health (ISSWSH) Expert Consensus Panel Review. Mayo Clin Proc 2017 Jan;92(1):114-128 [FREE Full text] [doi: 10.1016/j.mayocp.2016.09.018] [Medline: 27916394]
10. Stahl SM. Targeting circuits of sexual desire as a treatment strategy for hypoactive sexual desire disorder. J Clin Psychiatry 2010 Jul;71(7):821-822 [FREE Full text] [doi: 10.4088/JCP.10bs06117blu] [Medline: 20667289]
11. Uittenbogaart M, Leclercq WK, Bonouvrie D, Romeijn MM, Luijten AA, Olde Damink SW, et al. Diet-Induced Alteration of Microbiota and Development of Obesity, Nonalcoholic Fatty Liver Disease, and Diabetes: Study Protocol of a Prospective Study. JMIR Res Protoc 2019 Jun 19;8(6):e11553 [FREE Full text] [doi: 10.2196/11553] [Medline: 31219051]
12. Kindleysides S, Kruger R, Douwes J, Tannock GW, Renall N, Slater J, et al. Predictors Linking Obesity and the Gut Microbiome (the PROMISE Study): Protocol and Recruitment Strategy for a Cross-Sectional Study on Pathways That Affect the Gut Microbiome and Its Impact on Obesity. JMIR Res Protoc 2019 Aug 26;8(8):e14529 [FREE Full text] [doi: 10.2196/14529] [Medline: 31452525]
13. Mayer EA, Tillisch K, Gupta A. Gut/brain axis and the microbiota. J Clin Invest 2015 Mar 02;125(3):926-938 [FREE Full text] [doi: 10.1172/JCI76304] [Medline: 25689247]
14. González-Arancibia C, Urrutia-Piñones J, Illanes-González J, Martinez-Pinto J, Sotomayor-Zárate R, Julio-Pieper M, et al. Do your gut microbes affect your brain dopamine? Psychopharmacology (Berl) 2019 May;236(5):1611-1622. [doi: 10.1007/s00213-019-05265-5] [Medline: 31098656]
15. Clarke G, Grenham S, Scully P, Fitzgerald P, Moloney RD, Shanahan F, et al. The microbiome-gut-brain axis during early life regulates the hippocampal serotonergic system in a sex-dependent manner. Mol Psychiatry 2013 Jun;18(6):666-673 [FREE Full text] [doi: 10.1038/mp.2012.77] [Medline: 22688187]
16. Ho PSY, Jackson S, Cao S, Kwok C. Sex With Chinese Characteristics: Sexuality Research in/on 21st-Century China. J Sex Res 2018;55(4-5):486-521. [doi: 10.1080/00224499.2018.1437593] [Medline: 29521527]
17. Russomanno J, Patterson JG, Jabson Tree JM. Social Media Recruitment of Marginalized, Hard-to-Reach Populations: Development of Recruitment and Monitoring Guidelines. JMIR Public Health Surveill 2019 Dec 02;5(4):e14886 [FREE Full text] [doi: 10.2196/14886] [Medline: 31789598]
18. Ibarra JL, Agas JM, Lee M, Pan JL, Buttenheim AM. Comparison of Online Survey Recruitment Platforms for Hard-to-Reach Pregnant Smoking Populations: Feasibility Study. JMIR Res Protoc 2018 Apr 16;7(4):e101 [FREE Full text] [doi: 10.2196/resprot.8071] [Medline: 29661751]
19. Burns J, Johnstone K, Chavanduka T, Jamison C, Pena V, Stephenson R, et al. Evaluation of the Sexual Health Behaviors of Black Male Adolescents and Young Adults Through Social Media Platforms: Web-Based Survey Study. JMIR Public Health Surveill 2020 Sep 22;6(3):e19219 [FREE Full text] [doi: 10.2196/19219] [Medline: 32693387]
20. Motoki Y, Miyagi E, Taguri M, Asai-Sato M, Enomoto T, Wark JD, et al. Comparison of Different Recruitment Methods for Sexual and Reproductive Health Research: Social Media-Based Versus Conventional Methods. J Med Internet Res 2017 Mar 10;19(3):e73 [FREE Full text] [doi: 10.2196/jmir.7048] [Medline: 28283466]
21. Duncan DT, Park SH, Hambrick HR, Dangerfield Ii DT, Goedel WC, Brewer R, et al. Characterizing Geosocial-Networking App Use Among Young Black Men Who Have Sex With Men: A Multi-City Cross-Sectional Survey in the Southern United States. JMIR Mhealth Uhealth 2018 Jun 14;6(6):e10316 [FREE Full text] [doi: 10.2196/10316] [Medline: 29903702]
22. Christensen T, Riis AH, Hatch EE, Wise LA, Nielsen MG, Rothman KJ, et al. Costs and Efficiency of Online and Offline Recruitment Methods: A Web-Based Cohort Study. J Med Internet Res 2017 Mar 01;19(3):e58 [FREE Full text] [doi: 10.2196/jmir.6716] [Medline: 28249833]
23. Honka AM, Helander E, Pavel M, Jimison H, Mustonen P, Korhonen I, et al. Exploring Associations Between the Self-Reported Values, Well-Being, and Health Behaviors of Finnish Citizens: Cross-Sectional Analysis of More Than 100,000 Web-Survey Responses. JMIR Ment Health 2019 Apr 22;6(4):e12170 [FREE Full text] [doi: 10.2196/12170] [Medline: 31008710]

24. Savolainen I, Oksanen A, Kaakinen M, Sirola A, Paek H. The Role of Perceived Loneliness in Youth Addictive Behaviors: Cross-National Survey Study. JMIR Ment Health 2020 Jan 02;7(1):e14035 [FREE Full text] [doi: 10.2196/14035] [Medline: 31895044]
25. KEGG PATHWAY Database. 2021. URL: https://www.genome.jp/kegg/pathway.html [accessed 2021-02-03]
26. Messaoudi M, Lalonde R, Violle N, Javelot H, Desor D, Nejdi A, et al. Assessment of psychotropic-like properties of a probiotic formulation (Lactobacillus helveticus R0052 and Bifidobacterium longum R0175) in rats and human subjects. Br J Nutr 2011 Mar;105(5):755-764. [doi: 10.1017/S0007114510004319] [Medline: 20974015]
27. Steenbergen L, Sellaro R, van Hemert S, Bosch JA, Colzato LS. A randomized controlled trial to test the effect of multispecies probiotics on cognitive reactivity to sad mood. Brain Behav Immun 2015 Aug;48:258-264 [FREE Full text] [doi: 10.1016/j.bbi.2015.04.003] [Medline: 25862297]
28. Bancroft J, Janssen E, Strong D, Carnes L, Vukadinovic Z, Long JS. The relation between mood and sexuality in heterosexual men. Arch Sex Behav 2003 Jun;32(3):217-230. [doi: 10.1023/a:1023409516739] [Medline: 12807294]
29. Diaz Heijtz R, Wang S, Anuar F, Qian Y, Björkholm B, Samuelsson A, et al. Normal gut microbiota modulates brain development and behavior. Proc Natl Acad Sci U S A 2011 Feb 15;108(7):3047-3052 [FREE Full text] [doi: 10.1073/pnas.1010529108] [Medline: 21282636]
30. Cryan JF, Dinan TG. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. Nat Rev Neurosci 2012 Oct;13(10):701-712. [doi: 10.1038/nrn3346] [Medline: 22968153]
31. Desbonnet L, Garrett L, Clarke G, Bienenstock J, Dinan TG. The probiotic Bifidobacteria infantis: An assessment of potential antidepressant properties in the rat. J Psychiatr Res 2008 Dec;43(2):164-174. [doi: 10.1016/j.jpsychires.2008.03.009] [Medline: 18456279]
32. Pokusaeva K, Johnson C, Luk B, Uribe G, Fu Y, Oezguen N, et al. GABA-producing Bifidobacterium dentium modulates visceral sensitivity in the intestine. Neurogastroenterol Motil 2017 Jan;29(1):1 [FREE Full text] [doi: 10.1111/nmo.12904] [Medline: 27458085]
33. Puertollano E, Kolida S, Yaqoob P. Biological significance of short-chain fatty acid metabolism by the intestinal microbiome. Curr Opin Clin Nutr Metab Care 2014 Mar;17(2):139-144. [doi: 10.1097/MCO.0000000000000025] [Medline: 24389673]
34. Kratsman N, Getselter D, Elliott E. Sodium butyrate attenuates social behavior deficits and modifies the transcription of inhibitory/excitatory genes in the frontal cortex of an autism model. Neuropharmacology 2016 Mar;102:136-145. [doi: 10.1016/j.neuropharm.2015.11.003] [Medline: 26577018]
35. Weir TL, Manter DK, Sheflin AM, Barnett BA, Heuberger AL, Ryan EP. Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. PLoS One 2013;8(8):e70803 [FREE Full text] [doi: 10.1371/journal.pone.0070803] [Medline: 23940645]
36. Varfi N, Rothen S, Jasiowka K, Lepers T, Bianchi-Demicheli F, Khazaal Y. Sexual Desire, Mood, Attachment Style, Impulsivity, and Self-Esteem as Predictive Factors for Addictive Cybersex. JMIR Ment Health 2019 Jan 21;6(1):e9978 [FREE Full text] [doi: 10.2196/mental.9978] [Medline: 30664470]
37. Edmonds MR, Hadjistavropoulos HD, Gullickson KM, Asmundson AJ, Dear BF, Titov N. Understanding Problems With Sleep, Sexual Functioning, Energy, and Appetite Among Patients Who Access Transdiagnostic Internet-Delivered Cognitive Behavioral Therapy for Anxiety and Depression: Qualitative Exploratory Study. JMIR Form Res 2020 Oct 13;4(10):e15037 [FREE Full text] [doi: 10.2196/15037] [Medline: 33048054]
38. He Y, Wu W, Zheng H, Li P, McDonald D, Sheng H, et al. Regional variation limits applications of healthy gut microbiome reference ranges and disease models. Nat Med 2018 Oct;24(10):1532-1535. [doi: 10.1038/s41591-018-0164-x] [Medline: 30150716]
39. Graziottin A. Prevalence and evaluation of sexual health problems--HSDD in Europe. J Sex Med 2007 Mar;4 Suppl 3:211-219. [doi: 10.1111/j.1743-6109.2007.00447.x] [Medline: 17394593]

Abbreviations

GABA: γ-aminobutyric acid
HSDD: hypoactive sexual desire disorder
NHSD: no history of sexual dysfunction
SCFA: short-chain fatty acids

Edited by R Kukafka; submitted 28.10.20; peer-reviewed by B Ma, S Yang; comments to author 10.11.20; accepted 20.01.21; published 25.02.21.

Please cite as:
Li G, Li W, Song B, Wang C, Shen Q, Li B, Tang D, Xu C, Geng H, Gao Y, Wang G, Wu H, Zhang Z, Xu X, Zhou P, Wei Z, He X, Cao Y
Differences in the Gut Microbiome of Women With and Without Hypoactive Sexual Desire Disorder: Case Control Study
J Med Internet Res 2021;23(2):e25342
URL: https://www.jmir.org/2021/2/e25342
doi:10.2196/25342
PMID:33629964

Original Paper

Online Mental Health Animations for Young People: Qualitative Empirical Thematic Analysis and Knowledge Transfer

Helen Coughlan[1], BSS, MPhil, PhD; David Quin[2], BA, MARes; Kevin O'Brien[3], BA, MSc; Colm Healy[1], BSc, MSc, PhD; Jack Deacon[3*], BSc; Naoise Kavanagh[4*], MSc; Niamh Humphries[5], BSocSci, MPhil, PhD; Mary C Clarke[6], BA, MSc, PhD; Mary Cannon[1], MB, BCh, BAO, MSc, PhD

[1]Royal College of Surgeons in Ireland, Department of Psychiatry, Dublin, Ireland

[2]Department of Film and Media, Institute of Art, Design and Technology, Dun Laoghaire, Ireland

[3]SpunOut.ie, Dublin, Ireland

[4]Jigsaw, The National Centre for Youth Mental Health, Dublin, Ireland

[5]Royal College of Physicians of Ireland, Dublin, Ireland

[6]Department of Psychology, Royal College of Surgeons in Ireland, Dublin, Ireland

[*]these authors contributed equally

Corresponding Author:
Helen Coughlan, BSS, MPhil, PhD
Royal College of Surgeons in Ireland
Department of Psychiatry
RCSI Education & Research Centre
Dublin 9, D09 YD60
Ireland
Phone: 353 1 8093855
Email: helencoughlan@rcsi.ie

Abstract

Background: Mental ill-health is one of the most significant health and social issues affecting young people globally. To address the mental health crisis, a number of cross-sectoral research and action priorities have been identified. These include improving mental health literacy, translating research findings into accessible public health outputs, and the use of digital technologies. There are, however, few examples of public health–oriented knowledge transfer activities involving collaborations between researchers, the Arts, and online platforms in the field of youth mental health.

Objective: The primary aim of this project was to translate qualitative research findings into a series of online public mental health animations targeting young people between the ages of 16 and 25 years. A further aim was to track online social media engagement and viewing data for the animations for a period of 12 months.

Methods: Qualitative data were collected from a sample of 17 youth in Ireland, aged 18-21 years, as part of the longitudinal population-based Adolescent Brain Development study. Interviews explored the life histories and the emotional and mental health of participants. The narrative analysis revealed 5 thematic findings relating to young people's emotional and mental health. Through a collaboration between research, the Arts, and the online sector, the empirical thematic findings were translated into 5 public health animations. The animations were hosted and promoted on 3 social media platforms of the Irish youth health website called SpunOut. Viewing data, collected over a 12-month period, were analyzed to determine the reach of the animations.

Results: Narrative thematic analysis identified anxiety, depression, feeling different, loneliness, and being bullied as common experiences for young people. These thematic findings formed the basis of the animations. During the 12 months following the launch of the animations, they were viewed 15,848 times. A majority of views occurred during the period of the social media ad campaign at a cost of €0.035 (approximately US $0.042) per view. Animations on feeling different and being bullied accounted for the majority of views.

Conclusions: This project demonstrates that online animations provide an accessible means of translating empirical research findings into meaningful public health outputs. They offer a cost-effective way to provide targeted online information about mental health, coping, and help-seeking to young people. Cross-sectoral collaboration is required to leverage the knowledge and expertise required to maximize the quality and potential reach of any knowledge transfer activities. A high level of engagement

is possible by targeting non–help-seeking young people on their native social media platforms. Paid promotion is, therefore, an important consideration when budgeting for online knowledge translation and dissemination activities in health research.

(J Med Internet Res 2021;23(2):e21338) doi:10.2196/21338

KEYWORDS

mental health; public health; mental health literacy; social media; youth; qualitative; knowledge translation; anxiety; bullying; depression; loneliness; internet

Introduction

> *If health is the goal, biomedical interventions are not the only means to it. A broadened perspective expands the range of health-promoting practices and enlists the collective efforts of researchers and practitioners who have much to contribute from a variety of disciplines to the health of a nation [1].*

Mental health is one of the most significant health and social issues affecting young people globally [2-4]. Mental disorders are the leading cause of disability for people aged 10-24 years [5]. Developing mental ill-health during youth places young people at risk of enduring mental health difficulties [6], which are accompanied by a myriad of social, vocational, and relational consequences [2,3]. Thus, promotion, prevention, and early intervention for young people who may be at risk of developing mental ill-health are global health imperatives [7-12]. The Lancet Commission on Global Mental Health and Sustainable Development [13] has recommended a broad range of approaches across multiple sectors to address the global mental health crisis. Included in these are health promotion and the need to hear from those who have experienced mental health difficulties.

Lack of knowledge about features and signs of mental health difficulties (mental health literacy) and how to access support are both associated with mental health treatment avoidance or delay [14]. However, public health campaigns have been shown to be effective in changing both attitudes and intended behavior, including help-seeking [15-18]. Mental health campaigns that promote mental health literacy, personalize and normalize the experience of mental health difficulties, and have a recovery orientation have been found to both reduce stigma and promote help-seeking behavior [15,18-20]. The effectiveness of public health campaigns can be enhanced by ensuring that messages are well-designed and target and reach intended audiences [16]. In the field of youth mental health, the use of digital technologies and web-based platforms has been identified as an essential way of reaching young people and delivering both mental health information and support [13,21]. At least 91% of Europeans aged 16 to 29 years use the internet on a daily basis [22], and evidence suggests that young people are turning to web-based platforms to access health and mental health information and advice [23-26]. Findings from a recent survey of over 19,000 Irish youth suggested that, after family and friends, the internet is where 20% of adolescents and 33% of young adults go to informally seek information or support on mental health [24]. The anonymity, ease of access, absence of financial or educational barriers, and the nonstigmatizing environment offered by web-based mental health platforms have been identified as positive features of web-based mental health information by young people [27]. Thus, the internet is an ideal space for public health knowledge transfer outputs.

In their study, Wetterlin et al [23] found that 72.3% of respondents aged 17-24 years rated access to videos explaining mental health issues as highly important on web-based platforms. Among the many multimedia formats that can be used, animations have particular potential for public health communication [28]. They have the potential to provide strong symbolic representation of concepts. Additionally, as they are often short in length, they are considered to be an efficient way to communicate complex issues succinctly, to promote learning [29], and to influence intentions to change health-focused behavior [30]. Importantly, they also offer the potential to communicate health information across all levels of literacy [31,32]. This is particularly the case for spoken animations, which have been found to be the most effective way to communicate complex health information to people with low literacy levels [28,31]. In their study, George et al [32] found that people's responses to video animations were overwhelmingly positive, with most perceiving animation to be more engaging and relatable than other information video formats. Pacing, tone, and character rendering were rated as important factors in individuals' responses to animations.

Although increasingly a requirement of health research funders [33], there is a dearth of published material documenting knowledge transfer activities in the field of youth mental health research. In this paper, we describe a collaborative knowledge transfer project involving the translation of qualitative research findings on young people's emotional and mental health into online public health animations.

The project was conceived in response to a Knowledge Exchange and Dissemination Scheme funding call from the Health Research Board in Ireland. The scheme supports dissemination activities aimed at the general public or specific subgroups of the general public and is open to existing Health Research Board grant-holders. HC, the lead author, was conducting qualitative research on the lives and mental health of young people as part of a Health Research Board grant-funded PhD. Emergent findings from her research had provided compelling insights into young people's lived experiences of emotional and mental health struggles. HC recognized that the reach and impact of her research could be increased significantly if the findings could be meaningfully and creatively translated and shared with other young people. This resulted in a successful application for the Youth Mental Health Animation Creation Project by HC.

The aim of the project was to develop engaging and accessible public mental health animations for young people. The project involved a collaboration between research (Royal College of Surgeons in Ireland), the Arts (the Institute of Art, Design, and Technology), and the online youth health sector (SpunOut). The IADT Animation department was invited to join as a project partner because of its previous experience in translating complex and emotive material through animation for health and mental health organizations in Ireland. To maximize the potential reach of the animations, Ireland's leading health information website for young people aged 16 to 25 years, SpunOut, also joined as a project partner. SpunOut has over 1.2 million unique users per year with an average of 180,000 individuals accessing content per month [34]. The project was conducted in phases: research data collection and analysis; developing narrative scripts using qualitative data; creating and promoting the animations; and collection and analysis of online engagement and view data.

Methods

Study Population

Qualitative data were collected from 17 young people (10 male, 7 female) aged 18-21 years from the Adolescent Brain Development study [35,36], a longitudinal, epidemiological, population-based study that has been examining mental health and brain development among Irish youth since 2007. At the time of the animation project, 3 waves of data collection had been completed: (1) a baseline clinical interview study of 211 young people aged 11-13 years; (2) a follow-up clinical interview study of 86 individuals aged 14-18 years; and (3) a nested qualitative follow up study with a subsample of 17 individuals aged 18-21 years. The aim of the qualitative study was to explore young people's life narratives with a focus on adverse life experiences, interpersonal relationships, mental health and subjective well-being. Findings from the 17 individuals who took part in the qualitative study at follow up 2 formed the basis of the animations.

Data Collection

For the qualitative study, data were collected using in-depth qualitative interviews. These were conducted by HC from May 9 to July 25, 2016. Interviews lasted between 45 minutes and 1 hour 50 minutes. A semistructured interview schedule was used to explore participants' early family life experiences, their experiences of adverse or stressful life events, their mental health, their subjective well-being, their relationships with family and peers, their self-perception, their educational and vocational experiences, and their satisfaction with life. These were explored over each individual's life course. Written consent, which included consent to audio record study interviews was obtained from all participants. Participants were compensated for their time with a gift voucher. Audio recordings were transcribed by an external transcription agency and were subject to a nondisclosure agreement. All transcripts were subsequently checked for accuracy by HC.

Ethical approval was granted by the Research Ethics Committee of the Royal College of Surgeons in Ireland (RCSI REC 1221 March 2016).

Data Analysis

Interview data were analyzed using narrative analysis. Narrative analysis refers to a suite of methods that focus on the interpretation of individuals' lives as told in storied form [37,38]. Narrative analysis recognizes that all knowledge is constructed through multiple subjective interpretations of an individual's lived experiences and involves a dynamic interplay of subjectivity, perception, meaning, and context involving both the individuals who tell their stories and the researchers who listen and interpret those stories [39]. Specifically, as noted by Kirkman [40], narrative theory offers researchers the ability to "both to retain the complexity of the individual lives they study and to investigate multiple interactions among individuals and cultures." As a method, it focuses on the stories people tell about their life experiences across time, each of which is understood to have specific meaning to the person telling their story [41].

For this study, thematic narrative analysis [38] was used to identify themes within and across individuals' life stories. Although some forms of narrative analysis focus on both story content and how people tell their stories, the exclusive focus of thematic narrative analysis is the content of the stories that people tell [38]. However, unlike other thematic methods, such as grounded theory, it focuses on maintaining the integrity of individuals' stories during the analysis rather than on extracting decontextualized themes across cases [38]. Drawing specifically on the work of McCormack [39], we initially analyzed the construction of interpretive life story summaries for each participant. This process involved repeated listening to and reading of the qualitative interviews, during which notes and memos were documented. Each individual's life story was then mapped visually (in a mind map format and sequentially, from birth to the time of interview) and a life history summary was written for each individual based on an interpretation of each life story as told by each individual and interpreted by HC. Life story summaries were then examined for key themes within each individual's life story. The analysis method used to identify themes for each individual was that described by Braun and Clarke [42,43]. It involved the generation of inductive thematic codes for each participant based on both the manifest and latent themes across their life stories. These codes were then examined and combined into broader descriptive themes, which included a number of themes relating to participants' emotional and mental health.

Once coding was completed for each individual, findings were compared across all participants to identify any shared themes across the sample as a whole. For the animation project, we focused only on thematic findings relating to the mental health of individuals during their midadolescent and early adult years. The rationale for this was that the midadolescent and early adult phase of the lifespan is a peak period of risk for the onset of mental health difficulties [6,44]. We wanted the animations to reflect mental health experiences reported by young people during this potentially vulnerable phase of their lives. It also fit with the 16-to-25-year-old target age range of SpunOut. Five dominant mental health themes were identified across participants' subjective accounts of issues relating to their emotional and mental health during their mid-adolescent and early adult years. These were Anxiety, Depression, Feeling

Different, Loneliness and Being Bullied. These formed the basis for each animation.

Developing Narrative Scripts Using Qualitative Data

With evidence that videos of no more than 2 minutes duration are optimal to maximize viewer attention and engagement [45], we aimed to create 5 animations of between 60 and 120 seconds each. Furthermore, to ensure that the animations reflected the study findings and captured the authentic voices of young people, this phase involved developing composite narrative scripts for each of the 5 animations using verbatim quotes from multiple participants' interview data.

All individuals who had attended for interview were recontacted about the animation project. Of the 17 participants who had been interviewed, 7 replied. The project was discussed with each and written consent was sought to use quotes from their interviews to create the scripts. All 7 consented. Interview transcripts for these individuals were examined and any relevant quotes pertaining to the 5 animation themes were extracted. In the small number of instances where relevant quotes were not contained in the interview data from these individuals, quotes from other participants from the study were extracted, edited, and modified for inclusion in the script. Linking phrases were also added by HC to optimize the flow and necessary messaging of each narrative script.

Each script was written as a first-person account following a similar narrative arc based on social cognitive theory [1]. From a social cognitive perspective, positive health behaviors and behavior change are only possible when individuals understand health behaviors, have a belief in their capacity to control their health behaviors, and hold expectations about the possible outcomes for their actions [1]. For example, Meyerowitz and Chaiken [46] found that public health communications that enhanced individuals' sense of self-efficacy to take action in relation to their own health behaviors were most effective. Each script begins by describing the experience and how it feels, including the emotional, cognitive, physical, social, and relational aspects of the experience. Following this, the script incorporates ambivalence on the part of the young person, capturing young people's struggles to accept their own suffering and reach out for support. This is in line with existing evidence of ambivalence reported in the literature [47,48]. Each animation then highlights different actions taken to respond to the theme of the script. These include talking to informal supports, engaging in hobbies or other activities, speaking to a trusted adult, and accessing formal counselling and mental health supports. These actions were all reported by participants in the study. They also complement existing evidence on the protective roles of formal and informal supports for young people [49,50], trusted adults in a young person's life [51], and involvement in meaningful hobbies and social activities [52-54]. Each animation ends with a message that combines hope and realism. Specifically, that the action taken has enhanced the young person's sense of well-being and connectedness but that attending to emotional and mental health issues is an ongoing process and no single action is a panacea to the existential realities and challenges of the human experience [55]. Thus, in line with evidence on maximizing effectiveness in public health campaigns, the content of each animation incorporated information on mental health literacy and help-seeking, using first-person accounts with a recovery orientation [15,16,18,20]. Scripts (with associated subthemes) can be found in Multimedia Appendix 1.

Once each script had been crafted, all scripts were sent to the research participants who had consented to the use of quotes from their interviews. Scripts were marked for each individual to clarify which quotes had been extracted from their interview data. This was to offer participants an opportunity to withdraw their consent or to remove any quotes if they had any concerns about their anonymity. All participants were satisfied with the scripts as written.

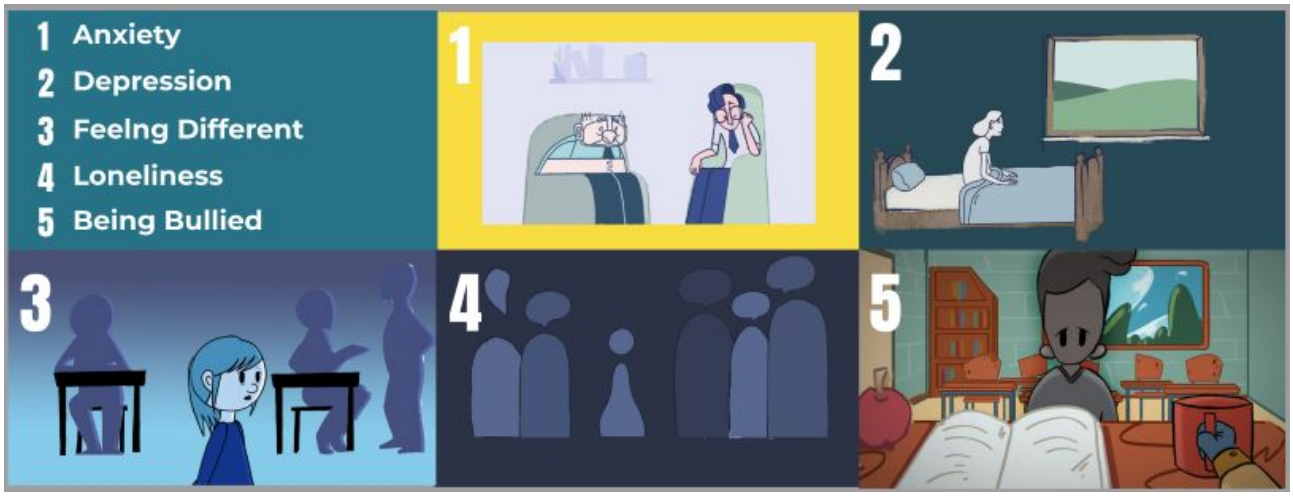
Animations

Creation

Animations were created in collaboration with the Animation program in the Film and Media Department of the IADT in Dublin. IADT is the only institute of art, design, and technology in Ireland that focuses specifically on the creative cultural and technological sectors. This animation project was integrated into the curriculum of third-year animation students in IADT as part of their applied professional practice learning. HC acted as executive producer and executive director for the animations, working with 5 student animation teams who crafted the scripts into the final animations. Students were overseen by DQ, the animation program lead.

The collaborative process was designed to maximize the authenticity and potential impact of each animation, while also protecting the research participants' data. The collaboration combined HC's expertise in mental health and the animation students' expertise in conceptualizing, creating and producing animations. It also enabled exploration and discussion of issues such as pacing, tone, and character rendering in each animation [32] (Figure 1). A decision was made by HC to ensure the design style was simple and minimalistic and that character rendering was not overly polished. This was to ensure that the style and rendering of the animations was as congruent as possible with the stories being told. Additionally, in line with evidence on how to maximize animation effectiveness [28,31], first-person narrative voice-overs of the scripts were layered onto the animations. Three of the voice-overs used female actors (depression, loneliness, and feeling different), and 2 used male voice-over actors (anxiety and bullying). The process of cocreating the animations lasted for approximately 4 months.

Figure 1. Screenshots from each of the 5 animations.



To promote accessibility, subtitled versions were developed for all animations. Subtitles are essential for individuals who are Deaf/deaf [56] and have also been found to enhance multimedia animation learning in people with attention deficit hyperactivity disorder [57]. Adding subtitles also ensured that the animations could be watched without audio, something of particular relevance to young people's use of mobile technology. Furthermore, although over 70,000 of people in Ireland speak Irish on a daily basis [58], there is an absence of youth mental health information in the vernacular of young native Irish speakers. To address this deficit, Irish language versions were also developed with support from *Conradh na Gaeilge* [59], a social and cultural organization that promotes the Irish language in Ireland and worldwide.

Once the animations had been completed, they were shown to the research participants who had consented to their verbatim quotes being used. This was to do a final check with those participants that they were satisfied with the animations and to affirm their consent for them to be hosted online. All participants reported being highly satisfied with the completed animations and consented to the launch phase of the project.

Promotion

A multimethod approach was adopted in relation to hosting and promoting the animations. First, new content was developed for SpunOut connected to each of the animations. This new content was embedded into the SpunOut website. All existing SpunOut content was also reviewed to identify relevant sections of the website where young people could access further information. Animations were hosted on a unique webpage [60]. Hyperlinks to this hosting page were included in all social media posts to facilitate young people who wished to view more of the animations. All 3 versions of the animations (nonsubtitled, subtitled, and Irish language versions) were also hosted on the SpunOut YouTube channel.

A launch event took place on May 9, 2019, using the hashtag #YMHanimate. Following this, SpunOut engaged in a social media advertising promotion campaign on both Facebook (€300.00, approximately US $362.21) and Twitter (€300.00, approximately US $362.21). The target demographic for the promotion campaign was young people aged 16 to 25 years.

Collection and Analysis of Online Engagement and View Data

To determine online engagement, analytics data from SpunOut Twitter, Facebook, and YouTube accounts were collected and analyzed for the 12-month period following the launch event. A cut-off view length of 75% or above was used to determine viewing figures across all 3 platforms. One reason for choosing this view length was that analytics data on viewing figures can include views of as little as 2 seconds in length, rendering counts of "any views" unreliable. Additionally, the animation credits accounted for between 9% and 15% of the total view time of each animation. This meant that individuals did not have to watch the full duration of each animation to be exposed to the full content. Data on link clicks (where an individual clicked a related SpunOut content link after watching an animation on social media), costs per view, and viewer demographics could only be extracted from Facebook analytics. Available gender variables were restricted to male, female, and unknown.

Results

Views and Link Clicks

Over the period from May 9, 2019 to May 8, 2020, the animations were viewed by 15,848 young people across all social media platforms (based on our criterion of >75% view length). Facebook views accounted for almost two-thirds of all views. Feeling Different was the most viewed animation, followed by Being Bullied and Depression (see Table 1). A majority of views occurred over a period of approximately 2 months following the launch during the social media ad campaign. There were low rates of link clicks on Facebook (ie, when an individual clicked a link to content hosted on the main SpunOut webpage) with just 240 recorded during the period of the social media ad campaign. Further details on impressions and viewing figures are available in Table 1.

Table 1. Viewing figures across SpunOut Facebook, Twitter, and YouTube platforms.

Theme	Platform, n[a]			
	Facebook[b]	Twitter[b]	YouTube	Total
Anxiety	1119	594	1133	2846
Depression	2499	104	683	3286
Feeling Different	2447	1005	625	4077
Loneliness	990	81	519	1590
Being Bullied	3382	165	502	4049
Total	10,437	1949	3462	15,848

[a]The number of online users who viewed the animation for a minimum of 75% of its full length.

[b]Figures cover the period from May 9, 2019 to May 8, 2020. Most Facebook and Twitter engagement occurred during the period of the social media campaign.

Cost Per View and Reach

The cost per view of the animations on Facebook was €0.035 (approximately US $0.042) per young person. This was calculated based on the >75% view length count of 10,437. The cost per reach was €0.003 (approximately US $0.004) per young person, based on a total Facebook reach of 118,142.

Demographics

Available data from Facebook revealed that 55.1% of those (5750/10,437) who viewed the animations on Facebook were aged 18 to 24 years, 39.9% (4160/10,437) were aged 13 to 17 years, and the remaining 5.0% (527/10,437) were aged 25 years or over. There were higher rates of female viewers than males across all animations and age ranges with the exception of the Being Bullied animation, where higher rates of male views were observed across all age ranges (see Table 2)

Table 2. Age and gender of viewers on the SpunOut Facebook platforms from May 9, 2019 to May 8, 2020.

Theme	Age range			
	13-17 years (n=4160), n (%)	18-24 years (n=5750), n (%)	25+ years (n=527), n (%)	All (N=10,437), n (%)
Anxiety				
Total	516 (46.1)	506 (45.2)	97 (8.7)	1119 (100)
Male	160 (31.0)	213 (42.1)	39 (40.2)	412 (36.8)
Female	353 (68.4)	290 (57.3)	58 (59.8)	701 (62.6)
Unknown	3 (0.6)	3 (0.6)	0 (0.0)	6 (0.6)
Depression				
Total	988 (40.3)	1414 (57.7)	97 (4.0)	2499 (100)
Male	267 (27.0)	481 (34.0)	38 (39.2)	786 (31.5)
Female	719 (72.8)	929 (65.7)	59 (60.8)	1707 (68.3)
Unknown	2 (0.2)	4 (0.3)	0 (0.0)	6 (0.2)
Feeling different				
Total	869 (35.5)	1451 (59.3)	127 (5.2)	2447 (100)
Male	186 (21.4)	341 (23.5)	31 (24.4)	558 (22.8)
Female	674 (77.6)	1099 (75.7)	96 (75.6)	1869 (76.4)
Unknown	9 (1.0)	11 (0.8)	0 (0.0)	20 (0.8)
Loneliness				
Total	429 (43.3)	534 (53.9)	27 (2.7)	990 (100)
Male	157 (36.6)	168 (31.5)	13 (48.1)	338 (34.1)
Female	270 (62.9)	361 (67.6)	14 (51.9)	645 (65.2)
Unknown	2 (0.5)	5 (0.9)	0 (0.0)	7 (0.7)
Being bullied				
Total	1358 (40.2)	1845 (54.6)	179 (5.3)	3382 (100)
Male	687 (50.6)	1105 (59.9)	100 (55.9)	1892 (55.9)
Female	664 (48.9)	733 (39.7)	78 (43.6)	1475 (43.7)
Unknown	7 (0.5)	7 (0.4)	1 (0.6)	15 (0.4)

Discussion

General

This is the first knowledge transfer project we are aware of that has translated qualitative research findings on issues affecting young people's emotional and mental health into a series of bilingual public health online animations. In line with recent recommendations on addressing the global mental health crisis [13,21], the project has given voice to the lived experiences of young people who are struggling with their mental health using a collaborative knowledge transfer process. Our research revealed that experiences of anxiety, depression, loneliness, feeling different, and being bullied were common in the lives of young people during their midadolescent and early adult years. These findings were successfully translated into 5 public health animations through a unique collaboration between the research, Arts, and online sectors. All 5 animations were hosted and promoted by SpunOut [60]. In the 12 months following the launch of the animations, high engagement and viewing numbers were evident across SpunOut social media platforms for all 5 animations, with close to 16,000 views. A majority of engagement occurred during the limited period of the social media ad campaign. The animations exploring Feeling Different and Being Bullied had the highest number of views.

Comparison With Existing Research and Knowledge

Our qualitative research findings, highlighting young people's lived experiences of anxiety, depression, loneliness, feeling different, and bullying, are aligned to existing evidence. Epidemiological evidence in Ireland has found that, by the age of 24 years, over 1-in-4 young people in Ireland will have experienced clinical levels of anxiety (26.7%) and depression (28.5%) [61]. More recently, in their study of over 19,000 adolescents and young adults in Ireland, Dooley et al [24] found that 49% of adolescents and 58% of young adults were experiencing anxiety and 40% of adolescents and 58% of young adults were experiencing depression [24]. Anxiety and depression in youth populations have also been recently identified as a significant health issue internationally [62-64].

Our findings on loneliness and feeling different during the adolescent and early adult years complement existing evidence that youth is a key period of risk for loneliness and social disconnection [65-67]. Not only did this emerge as a qualitative theme in the research study, but the Feeling Different animation had the highest number of views. Loneliness and a sense of feeling different are associated with individuals' needs to explore and find their own identities during adolescence and early adulthood [65]. However, other factors such as culture, environment, personality factors, and gender are also implicated in experiences of loneliness [67]. In the case of gender, females are more likely to report feelings of loneliness and social disconnection, supporting our use of a female voiceover for the loneliness and feeling different animations. The female character representations used in these animations may also have been more congruent with the experiences of females, as reflected in the high prevalence of female views of these animations.

Our finding that many participants in the study had experienced bullying and the high view rate of the Being Bullied animation is consistent with recent Irish data showing rates of 39% and 58% among adolescents and young adults respectively who reported being the victim of bullying [24]. Rates among adolescents are similar to those reported internationally. In their meta-analysis on bullying, Modecki et al [68] found a mean prevalence rate of 35% for traditional bullying and 15% for cyberbullying across the 80 studies in their review. Additionally, our finding that higher numbers of males aged 18 years or older viewed the Being Bullied animation reflects gender trends in the national My World survey [24] of Irish youth where rates of bullying in males increased over time. Specifically, fewer males than females reported being bullied during adolescence (male: 40%, female: 45%) but a higher proportion reported being bullied during their young adult years (male: 61%, female: 57%).

The finding that the animations were viewed almost 16,000 times following their launch demonstrates the potential reach that animations can have within the youth mental health arena. To achieve this, a low budget social media ad campaign was required, and a majority of views occurred in response to this campaign over the campaign period of approximately 2 months. This highlights the value of animations as a medium for knowledge transfer [28,29,31,32] and the importance of budgeting for paid social media promotion to maximize the reach of multimedia knowledge transfer outputs. When proactively seeking mental health information, evidence suggests that young people prefer seeking information from information-based rather than social media websites [23,27,69]. However, for young people who are not proactively seeking mental health information, our high viewing figures during the promotion campaign indicate that social media campaigns may be a particularly effective method to engage non–help-seeking young people with public mental health information. Key to this is following social media trends in order to target those web-based and social media platforms that young people are already using [70].

We anticipated that the animations would enhance mental health literacy in young people, promote disclosure of mental health difficulties, and ensure young people understood how to access informal and formal mental health support. This was based on existing evidence that has shown that public awareness mental health campaigns are effective in achieving both attitude and behavior changes [14,15,18]. For example, an evaluation into the Time-to-Change public mental health awareness campaign in the United Kingdom found that individuals who were simply aware of the campaign reported increased comfort in disclosing mental health difficulties to family and friends and were more likely to seek professional help [18]. Similarly, in their review, Kauer et al [69] found that an increase in mental health literacy was a facilitator of help-seeking among young people accessing information online. A review by Pretorius et al [27] also found that young people used online information and resources to facilitate personal coping responses or as a means to promote informal support seeking behaviors and that the process of help-seeking online could act as a gateway to further help-seeking by connecting young people with information and additional supports. Based on this existing evidence, it is reasonable to hypothesize that, for a proportion of young people, viewing and being exposed to the messaging within each animation will have positively impacted their attitudes toward mental health difficulties, their mental health literacy (for those with low levels of mental health literacy), their willingness to share any current or future mental health concerns, and their willingness to reach out for information and support if they need to.

A final and important finding from this project was the low cost–high yield relationship between what was spent on social media promotion and the level of user engagement and views of the animations. In their systematic review of the use of social networking sites for public health practice and research, Capurro and colleagues [70] highlight that social media and social networking sites offer researchers the fast, easy, and low-cost access to a range of populations, making them an ideal platform for conducting research. Our cost-per-view findings, at a cost of just €0.003 (approximately US $0.004) to reach a young person with one of the videos and €0.035 (approximately US $0.042) to have a young person in our target demographic view the video to completion highlights the potential that social media promotion can offer in supporting impactful knowledge transfer activities targeting known populations at a very low cost. Moreover, our use of this method addresses a key factor that has been identified in maximizing the potential for public health messaging to change behavior: ensuring that messages are delivered to their intended audience with sufficient reach [16]. Additionally, it facilitated access to demographic and engagement data, an oftentimes underused data resource in the field of health-related research [70].

Limitations

A key limitation in this knowledge transfer project is that, while we were able to identify our target demographic for our Facebook and Twitter promotion campaigns, such promotion activity also relies on algorithms and models that are controlled by each social media or networking site. Thus, our demographic findings relating to the age and gender of user engagement may reflect aspects of the advertising algorithms used rather than solely reflecting gender trends related to the animations. Additionally, the idea to develop the animations was a response

to the findings emerging from the research. This meant that the focus of the animation project was on ensuring the animations were embedded and accessible to young people as part of an existing and reputable youth mental health website rather than on collecting data on young people's responses to the animations or their impact on attitudes, health, or help-seeking. This limited our analysis to the reach of the animations. We were therefore unable to examine the impact of the animations on those 15,000 or more young people who watched them or to interpret the low link click rate in our analytics data. However, in relation to the latter, it is important to note that link-click data were only collected during the short period of the social media and promotion period. The limitations of the reach data in this study highlight the importance of integrating research to evaluate impact and effectiveness when designing public health campaigns such as this. Future research is needed to examine the impact of both outputs, such as our animations, and the effectiveness of targeting young people on social media platforms.

Conclusions

In line with recommendations for tackling the global mental health crisis [13,21], this knowledge transfer project provides an example of how the mental health research community can engage in meaningful knowledge transfer activities targeting young people on their native social media platforms. By adopting this type of knowledge transfer activity, researchers have the potential to use and translate their findings to make a tangible difference to both individual lives and to overall societal health, beyond what is possible within the confines of traditional dissemination arenas and institutions.

Acknowledgments

We acknowledge the following parties—IADT Animators: Sarah Benson, Joshua Brown, Charlotte Connolly, Jessie Desmond Schmidt, Conor Doyle, Hollie Hamill, Anna Hoare, Sabine Jansone, Rory Martin, Beren McCormack, Jack McHugh, Giorgia McKenna, Conor McNally, Daniel McNicholl, Saoirse Mitchell, Chloe O'Hara, Krizsha Perocho, Cassandra Sierevogel, Carylle Sanchez, Charles-Henri Sanson, Aisling Smith, Iobhar Stokes Rodriguez, Sarah Whyte, Key Woodlock; IADT Sound: Tim Nerney, Geoffrey Perrin; Other Support: Health Service Executive, *Conradh na Gaeilge*, Derek Chambers.

This project was funded by the Health Research Board through grant funding to HC (KEDS-2017-014). The original research was funded by the Health Research Board through grant funding to HC (HPF-2015-974). NH is funded by the Health Research Board (EIA-2017-022). MCC is funded through the Health Research Board (HRA-2015) and the Irish Research Council (COALESCE/2019/61). MC and CH are funded through the European Research Council (724809).

Authors' Contributions

HC contributed to conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualization, and writing the original draft. DQ contributed to project administration, resources, supervision, validation, visualization, and reviewing and editing the manuscript. KO'B contributed to formal analysis, validation, and review and editing the manuscript. CH contributed to reviewing and editing the manuscript. JD contributed to formal analysis, resources, project administration, and reviewing and editing the manuscript. NK contributed to conceptualization, funding acquisition, and reviewing and editing the manuscript. NH contributed to funding acquisition, supervision, and reviewing and editing the manuscript. MCC contributed to funding acquisition, supervision, and reviewing and editing the manuscript. MC contributed to funding acquisition, supervision, and reviewing and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Animation themes, scripts, subthemes, health-promoting issues, and actions.
[PDF File (Adobe PDF File), 511 KB - jmir_v23i2e21338_app1.pdf]

References

1. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
2. McGorry PD, Purcell R, Hickie IB, Jorm AF. Investing in youth mental health is a best buy. Med J Aust 2007 Oct;187(S7):S5-S7. [doi: 10.5694/j.1326-5377.2007.tb01326.x]
3. Patel V, Flisher A, Hetrick S, McGorry P. Mental health of young people: a global public-health challenge. Lancet 2007 Apr 14;369(9569):1302-1313. [doi: 10.1016/S0140-6736(07)60368-7] [Medline: 17434406]
4. Coughlan H, Cannon M, Shiers D, Power P, Barry C, Bates T, et al. Towards a new paradigm of care: the International Declaration on Youth Mental Health. Early Interv Psychiatry 2013 May;7(2):103-108. [doi: 10.1111/eip.12048] [Medline: 23617573]

5. Mokdad A, Forouzanfar M, Daoud F, Mokdad A, El Bcheraoui C, Moradi-Lakeh M, et al. Global burden of diseases, injuries, and risk factors for young people's health during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2016 Jun 11;387(10036):2383-2401. [doi: 10.1016/S0140-6736(16)00648-6] [Medline: 27174305]
6. Kessler R, Amminger G, Aguilar-Gaxiola D, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry 2007 Jul;20(4):359-364 [FREE Full text] [doi: 10.1097/YCO.0b013e32816ebc8c] [Medline: 17551351]
7. McGorry P, Killackey E, Yung A. Early intervention in psychosis: concepts, evidence and future directions. World Psychiatry 2008 Oct;7(3):148-156 [FREE Full text] [doi: 10.1002/j.2051-5545.2008.tb00182.x] [Medline: 18836582]
8. McGorry P, Purcell R, Hickie I, Yung A, Pantelis C, Jackson HJ. Clinical staging: a heuristic model for psychiatry and youth mental health. Med J Aust 2007 Oct 01;187(S7):S40-S42. [doi: 10.5694/j.1326-5377.2007.tb01335.x] [Medline: 17908024]
9. Killackey EM, Alvarez-Jimenez M. Psychosocial interventions for youth mental health. In: McGorry P, Hickie I, editors. Clinical Staging in Psychiatry: Making Diagnosis Work for Research and Treatment. Cambridge: Cambridge University Press; 2019:241-260.
10. Wasserman D, Hoven C, Wasserman C, Wall M, Eisenberg R, Hadlaczky G, et al. School-based suicide prevention programmes: the SEYLE cluster-randomised, controlled trial. Lancet 2015 Apr 18;385(9977):1536-1544. [doi: 10.1016/S0140-6736(14)61213-7] [Medline: 25579833]
11. Malla A, Iyer S, McGorry P, Cannon M, Coughlan H, Singh S, et al. From early intervention in psychosis to youth mental health reform: a review of the evolution and transformation of mental health services for young people. Soc Psychiatry Psychiatr Epidemiol 2016 Mar;51(3):319-326. [doi: 10.1007/s00127-015-1165-4] [Medline: 26687237]
12. McGorry P, Nelson B. Why we need a transdiagnostic staging approach to emerging psychopathology, early diagnosis, and treatment. JAMA Psychiatry 2016 Mar;73(3):191-192. [doi: 10.1001/jamapsychiatry.2015.2868] [Medline: 26765254]
13. Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, Bolton P, et al. The Lancet Commission on global mental health and sustainable development. Lancet 2018 Oct 27;392(10157):1553-1598. [doi: 10.1016/S0140-6736(18)31612-X] [Medline: 30314863]
14. Henderson C, Evans-Lacko S, Thornicroft G. Mental illness stigma, help seeking, and public health programs. Am J Public Health 2013 May;103(5):777-780. [doi: 10.2105/AJPH.2012.301056] [Medline: 23488489]
15. Clement S, Jarrett M, Henderson C, Thornicroft G. Messages to use in population-level campaigns to reduce mental health-related stigma: consensus development study. Epidemiol Psichiatr Soc 2010;19(1):72-79. [doi: 10.1017/s1121189x00001627] [Medline: 20486426]
16. Abroms LC, Maibach EW. The effectiveness of mass communication to change public behavior. Annu Rev Public Health 2008;29:219-234. [doi: 10.1146/annurev.publhealth.29.020907.090824] [Medline: 18173391]
17. Snyder LM, Hamilton MA. A meta-analysis of U.S. health campaign effects on behavior: emphasize enforcement, exposure, and new information, and beware the secular trend. Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews. York: Centre for Reviews and Dissemination (UK); 2002. URL: https://www.ncbi.nlm.nih.gov/books/NBK69351/ [accessed 2021-01-26]
18. Henderson C, Robinson E, Evans-Lacko S, Thornicroft G. Relationships between anti-stigma programme awareness, disclosure comfort and intended help-seeking regarding a mental health problem. Br J Psychiatry 2017 Nov;211(5):316-322 [FREE Full text] [doi: 10.1192/bjp.bp.116.195867] [Medline: 28935661]
19. Jorm A, Korten A, Jacomb P, Christensen H, Rodgers B, Pollitt P. "Mental health literacy": A survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. Med J Aust 1997 Feb 17;166(4):182-186. [doi: 10.5694/j.1326-5377.1997.tb140071.x] [Medline: 9066546]
20. Pinfold V, Thornicroft G, Huxley P, Farmer P. Active ingredients in anti-stigma programmes in mental health. Int Rev Psychiatry 2005 Apr;17(2):123-131. [doi: 10.1080/09540260500073638] [Medline: 16194782]
21. Mei C, Fitzsimons J, Allen N, Alvarez-Jimenez M, Amminger GP, Browne V, et al. Global research priorities for youth mental health. Early Interv Psychiatry 2020 Feb;14(1):3-13. [doi: 10.1111/eip.12878] [Medline: 31960595]
22. Being young in Europe today. Eurostat. 2017. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Being_young_in_Europe_today [accessed 2021-01-26]
23. Wetterlin F, Mar M, Neilson E, Werker G, Krausz M. eMental health experiences and expectations: a survey of youths' web-based resource preferences in Canada. J Med Internet Res 2014 Dec 17;16(12):e293 [FREE Full text] [doi: 10.2196/jmir.3526] [Medline: 25519847]
24. Dooley B, O'Connor C, Fitzgerald A, O'Reilly A. My World Survey 2: national study of youth mental health in Ireland. Jigsaw, the National Centre for Youth Mental Health. 2019. URL: www.myworldsurvey.ie
25. Park E, Kwon M. Health-related internet use by children and adolescents: systematic review. J Med Internet Res 2018 Apr 03;20(4):e120 [FREE Full text] [doi: 10.2196/jmir.7731] [Medline: 29615385]
26. Atkinson N, Saperstein S, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009 Feb 20;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
27. Pretorius C, Chambers D, Cowan B, Coyle D. Young people seeking help online for mental health: cross-sectional survey study. JMIR Ment Health 2019 Aug 26;6(8):e13524 [FREE Full text] [doi: 10.2196/13524] [Medline: 31452519]

28. Mayer RR, Moreno R. Animation as an aid to multimedia learning. Edu Psychol Rev 2002 Mar;14:87-99. [doi: 10.1023/A:1013184611077]
29. Champoux JE. Comparative analyses of live - action and animated film remake scenes: finding alternative film-based teaching resources. Edu Media Int 2005 Mar;42(1):49-69. [doi: 10.1080/09523980500116662]
30. Lecky D, Dhillon H, Verlander N, McNulty C. Animations designed to raise patient awareness of prudent antibiotic use: patient recall of key messages and their immediate effect on patient attitude. BMC Res Notes 2017 Dec 06;10(1):701 [FREE Full text] [doi: 10.1186/s13104-017-3048-0] [Medline: 29208027]
31. Meppelink C, van Weert JCM, Haven C, Smit E. The effectiveness of health animations in audiences with different health literacy levels: an experimental study. J Med Internet Res 2015 Jan 13;17(1):e11 [FREE Full text] [doi: 10.2196/jmir.3979] [Medline: 25586711]
32. George S, Moran E, Duran N, Jenders R. Using animation as an information tool to advance health research literacy among minority participants. AMIA Annu Symp Proc 2013;2013:475-484 [FREE Full text] [Medline: 24551351]
33. Reed M, Bryce R, Machen R. Pathways to policy impact: a new approach for planning and evidencing research impact. Evid Policy 2018 Aug 31;14(3):431-458. [doi: 10.1332/174426418x15326967547242]
34. About us. Spunout. URL: https://spunout.ie/about [accessed 2020-05-08]
35. Coughlan H, Tiedt L, Clarke M, Kelleher I, Tabish J, Molloy C, et al. Prevalence of DSM-IV mental disorders, deliberate self-harm and suicidal ideation in early adolescence: an Irish population-based study. J Adolesc 2014 Jan;37(1):1-9. [doi: 10.1016/j.adolescence.2013.10.004] [Medline: 24331299]
36. Kelleher I, Murtagh A, Molloy C, Roddy S, Clarke M, Harley M, et al. Identification and characterization of prodromal risk syndromes in young adolescents in the community: a population-based clinical interview study. Schizophr Bull 2012 Mar;38(2):239-246 [FREE Full text] [doi: 10.1093/schbul/sbr164] [Medline: 22101962]
37. Bold C. Using Narrative in Research. London: Sage; 2011.
38. Riessman C. Narrative Methods for the Human Sciences. London: Sage; 2008.
39. Mccormack C. Storying stories: a narrative approach to in-depth interview conversations. Int J Soc Res Methodol 2004 Jul;7(3):219-236. [doi: 10.1080/13645570210166382]
40. Kirkman M. What's the plot? applying narrative theory to research in psychology. Aust Psychol 2002;37(1):30-38. [doi: 10.1080/00050060210001706646]
41. Stephens C. Narrative analysis in health psychology research: personal, dialogical and social stories of health. Health Psychol Rev 2011 Mar;5(1):62-78. [doi: 10.1080/17437199.2010.543385]
42. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
43. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
44. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
45. Fishman E. How long should your next video be? Wistia. 2016 Jul 05. URL: https://wistia.com/learn/marketing/optimal-video-length [accessed 2020-05-08] [WebCite Cache ID https://wistia.com/learn/marketing/optimal-video-length]
46. Meyerowitz BE, Chaiken S. The effect of message framing on breast self-examination attitudes, intentions, and behavior. J Pers Soc Psychol 1987 Mar;52(3):500-510. [doi: 10.1037//0022-3514.52.3.500] [Medline: 3572721]
47. Rickwood D, Deane FP, Wilson CJ, Ciarrochi J. Young people's help-seeking for mental health problems. Australian e-Journal for the Advancement of Mental Health 2014 Dec 17;4(3):218-251. [doi: 10.5172/jamh.4.3.218]
48. Rickwood DJ, Deane FP, Wilson CJ. When and how do young people seek professional help for mental health problems? Med J Aust 2007 Oct 01;187(S7):S35-S39. [doi: 10.5694/j.1326-5377.2007.tb01334.x] [Medline: 17908023]
49. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]
50. Barker G, Olukoya A, Aggleton P. Young people, social support and help-seeking. Int J Adolesc Med Health 2005;17(4):315-335. [doi: 10.1515/ijamh.2005.17.4.315] [Medline: 16445071]
51. Dooley B, Fitzgerald A, Giollabhui NM. The risk and protective factors associated with depression and anxiety in a national sample of Irish adolescents. Ir J Psychol Med 2015 Mar;32(1):93-105. [doi: 10.1017/ipm.2014.83] [Medline: 30185277]
52. Coughlan H, Healy C, Ní Sheaghdha A, Murray G, Humphries N, Clarke M, et al. Early risk and protective factors and young adult outcomes in a longitudinal sample of young people with a history of psychotic-like experiences. Early Interv Psychiatry 2020 Jun;14(3):307-320. [doi: 10.1111/eip.12855] [Medline: 31310453]
53. Werner E. Resilience and Recovery: findings from the Kauai longitudinal study. Research, Policy and Practice in Children's Mental Health 2005;19(1):11-14 [FREE Full text]
54. Werner E. Journeys from childhood to midlife: risk, resilience, and recovery. Pediatrics 2004 Aug;114(2):492. [doi: 10.1542/peds.114.2.492] [Medline: 15286237]
55. Van Deurrzen E. Existential therapy. In: Dryden W, editor. Dryden's Handbook of Individual Therapy. London: Sage; 2007:195-226.

56. Gernsbacher MA. Video captions benefit everyone. Policy Insights Behav Brain Sci 2015 Oct;2(1):195-202 [FREE Full text] [doi: 10.1177/2372732215602130] [Medline: 28066803]
57. Lewis D, Brown V. Multimedia and ADHD learners: are subtitles beneficial or detrimental? In: Proceedings of the Annual Convention of the Association for Educational Communications and Technology:Selected Research and Development Papers. 2012 Presented at: 35th Annual Convention of the Association for Educational Communications and Technology; October 3-November 3; Louisville, Kentucky p. 135-139 URL: https://files.eric.ed.gov/fulltext/ED546875.pdf#page=143
58. Census of population 2016 - Profile 10: education, skills and the Irish language. Central Statistics Office. URL: https://www.cso.ie/en/releasesandpublications/ep/p-cp10esil/p10esil/ilg/ [accessed 2020-05-19] [WebCite Cache ID http://www.cso.ie]
59. Conradh na Gaeilge. URL: https://www.cnag.ie/ga/ [accessed 2020-05-08]
60. SpunOut. URL: www.spunout.ie/RCSI [accessed 2020-05-08]
61. Cannon M, Coughlan H, Clarke M, Harley M, Kelleher I. The Mental Health of Young People in Ireland: A Report of the Psychiatric Epidemiology Research Across the Lifespan (PERL) Group. Dublin: Royal College of Surgeons in Ireland; Oct 2013.
62. Whitney DG, Shapiro DN, Warschausky SA, Hurvitz EA, Peterson MD. The contribution of neurologic disorders to the national prevalence of depression and anxiety problems among children and adolescents. Ann Epidemiol 2019 Jan;29:81-84.e2 [FREE Full text] [doi: 10.1016/j.annepidem.2018.11.003] [Medline: 30545763]
63. Spence S, Zubrick S, Lawrence D. A profile of social, separation and generalized anxiety disorders in an Australian nationally representative sample of children and adolescents: Prevalence, comorbidity and correlates. Aust N Z J Psychiatry 2018 May;52(5):446-460. [doi: 10.1177/0004867417741981] [Medline: 29185356]
64. Sandal R, Goel N, Sharma M, Bakshi R, Singh N, Kumar D. Prevalence of depression, anxiety and stress among school going adolescent in Chandigarh. J Family Med Prim Care 2017;6(2):405-410 [FREE Full text] [doi: 10.4103/2249-4863.219988] [Medline: 29302555]
65. Qualter P, Vanhalst J, Harris R, Van Roekel E, Lodder G, Bangee M, et al. Loneliness across the life span. Perspect Psychol Sci 2015 Mar;10(2):250-264. [doi: 10.1177/1745691615568999] [Medline: 25910393]
66. Lasgaard M, Friis K, Shevlin M. "Where are all the lonely people?" a population-based study of high-risk groups across the life span. Soc Psychiatry Psychiatr Epidemiol 2016 Oct;51(10):1373-1384. [doi: 10.1007/s00127-016-1279-3] [Medline: 27571769]
67. Mund M, Freuding M, Möbius K, Horn N, Neyer F. The stability and change of loneliness across the life span: A meta-analysis of longitudinal studies. Pers Soc Psychol Rev 2020 Feb;24(1):24-52 [FREE Full text] [doi: 10.1177/1088868319850738] [Medline: 31179872]
68. Modecki K, Minchin J, Harbaugh A, Guerra N, Runions K. Bullying prevalence across contexts: a meta-analysis measuring cyber and traditional bullying. J Adolesc Health 2014 Nov;55(5):602-611. [doi: 10.1016/j.jadohealth.2014.06.007] [Medline: 25168105]
69. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014 Mar 04;16(3):e66 [FREE Full text] [doi: 10.2196/jmir.3103] [Medline: 24594922]
70. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014 Mar 14;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]

Abbreviations

IADT: Institute of Art, Design and Technology

Edited by G Eysenbach, R Kukafka; submitted 12.06.20; peer-reviewed by D Rickwood, P Kreniske, S Singh; comments to author 18.08.20; revised version received 16.09.20; accepted 27.11.20; published 09.02.21.

Please cite as:
Coughlan H, Quin D, O'Brien K, Healy C, Deacon J, Kavanagh N, Humphries N, Clarke MC, Cannon M
Online Mental Health Animations for Young People: Qualitative Empirical Thematic Analysis and Knowledge Transfer
J Med Internet Res 2021;23(2):e21338
URL: http://www.jmir.org/2021/2/e21338/
doi:10.2196/21338
PMID:33560231

Original Paper

Differences in Basic Life Support Knowledge Between Junior Medical Students and Lay People: Web-Based Questionnaire Study

Ludovic Sturny[1], MSc; Simon Regard[1], MD, MPH; Robert Larribau[1], MD; Marc Niquille[1], MD; Georges Louis Savoldelli[2,3], MD, MEd; François Sarasin[1], MD; Eduardo Schiffer[2], MD; Laurent Suppan[1], MD

[1]Division of Emergency Medicine, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[2]Division of Anesthesiology, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[3]Unit of Development and Research in Medical Education, Faculty of Medicine, University of Geneva, Geneva, Switzerland

Corresponding Author:
Laurent Suppan, MD
Division of Emergency Medicine
Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine
University of Geneva Hospitals and Faculty of Medicine
Rue Gabrielle-Perret-Gentil 4
Geneva, 1211
Switzerland
Phone: 41 795532579
Email: laurent.suppan@hcuge.ch

Abstract

Background: Early cardiopulmonary resuscitation and prompt defibrillation markedly increase the survival rate in the event of out-of-hospital cardiac arrest (OHCA). As future health care professionals, medical students should be trained to efficiently manage an unexpectedly encountered OHCA.

Objective: Our aim was to assess basic life support (BLS) knowledge in junior medical students at the University of Geneva Faculty of Medicine (UGFM) and to compare it with that of the general population.

Methods: Junior UGFM students and lay people who had registered for BLS classes given by a Red Cross–affiliated center were sent invitation links to complete a web-based questionnaire. The primary outcome was the between-group difference in a 10-question score regarding cardiopulmonary resuscitation knowledge. Secondary outcomes were the differences in the rate of correct answers for each individual question, the level of self-assessed confidence in the ability to perform resuscitation, and a 6-question score, "essential BLS knowledge," which only contains key elements of the chain of survival. Continuous variables were first analyzed using the Student *t* test, then by multivariable linear regression. Fisher exact test was used for between-groups comparison of binary variables.

Results: The mean score was higher in medical students than in lay people for both the 10-question score (mean 5.8, SD 1.7 vs mean 4.2, SD 1.7; *P*<.001) and 6-question score (mean 3.0, SD 1.1 vs mean 2.0, SD 1.0; *P*<.001). Participants who were younger or already trained scored consistently better. Although the phone number of the emergency medical dispatch center was well known in both groups (medical students, 75/80, 94% vs lay people, 51/62, 82%; *P*=.06), most participants were unable to identify the criteria used to recognize OHCA, and almost none were able to correctly reorganize the BLS sequence. Medical students felt more confident than lay people in their ability to perform resuscitation (mean 4.7, SD 2.2 vs mean 3.1, SD 2.1; *P*<.001). Female gender and older age were associated with lower confidence, while participants who had already attended a BLS course prior to taking the questionnaire felt more confident.

Conclusions: Although junior medical students were more knowledgeable than lay people regarding BLS procedures, the proportion of correct answers was low in both groups, and changes in BLS education policy should be considered.

(J Med Internet Res 2021;23(2):e25125) doi:10.2196/25125

KEYWORDS

basic life support; cardiopulmonary resuscitation; medical students; undergraduate medical education; out-of-hospital cardiac arrest; life support; cardiopulmonary; medical education

Introduction

Basic life support (BLS) maneuvers and use of automated external defibrillators (AED) have been shown to greatly increase the survival rate after out-of-hospital cardiac arrests (OHCA) [1]. Nevertheless, their application rate remains very different from one region to another [2,3].

In Switzerland, most regions lack a systematic BLS training program for the general population, even though more than 8000 OHCA occur every year in the country [4]. In Geneva, BLS was provided in less than 40% of OHCA cases between 2009 and 2012 [5].

Medical students might unexpectedly encounter OHCA cases outside of the hospital or university environment and might, given their status, be expected to take care of the situation. Many studies carried out in different medical education systems around the world have however concluded that BLS knowledge among health care students is generally limited [6-10].

Since the first BLS training session for medical students of the University of Geneva – Faculty of Medicine (UGFM) only takes place during the second of their 6-year curriculum, our hypothesis was that they might lack critical knowledge regarding BLS prior to this course. These medical students might, however, be unpredictably faced with OHCA and be expected to respond swiftly and adequately given their chosen profession [11]. Our goal was to assess BLS knowledge among these students prior to their first BLS training and to compare it with that of the general population.

Methods

Study Design

A cross-sectional, web-based, questionnaire study compliant with the CHERRIES guidelines was carried out between October 2019 and April 2020 [12]. Such studies do not come within the scope of the Swiss Federal Act on Research involving Human Beings [13]. The study protocol was nevertheless submitted to the regional ethics committee, which declared the project “non-objectionable” (Request #2019-02047).

Online Platform and Survey Development

An internet platform was developed using the Joomla 3.9 content management system (Open Source Matters, New York, NY). The Community Surveys Pro component version 5 (CoreJoomla, Hyderabad, India) was used to create the questionnaire.

A structured online questionnaire containing 19 questions requiring either open or closed answers was created on the platform (Multimedia Appendix 1). The number of questions was kept below 20 to limit dropout attrition [14-16]. Ten questions were used to assess BLS knowledge. These questions were prepared according to the 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations [17]. The questionnaire was displayed over a total of 4 pages (Table 1). Both the internet platform and the questionnaire were thoroughly tested for usability and user-friendliness by all coauthors before beginning the study.

Table 1. Survey structure and questions.

Survey page and field, Question	Type of question
Introduction: consent	
N/A[a]	N/A
1: demographics	
Year of birth	Open (Regex[b])
Gender	MCQ[c]
2: general BLS[d] knowledge	
Ever heard of BLS or ACLS[e] before	Yes/No
Meaning of "AED"[f,g]	Open
Year of the last BLS guidelines update	Open (Regex)
Phone number of the emergency medical services dispatch center[g,h]	Open
3: prior BLS experience	
Current or past student of a health care profession, BLS instructor, or professional rescuer	MAQ[i]
Prior BLS training	Yes/No
Wish to be trained, or more trained, in BLS procedures	Yes/No
4: specific BLS knowledge	
Criteria used to recognize OHCA[g,h,j]	MAQ
BLS sequence[g,h]	Ordering
Artery for pulse assessment[g]	MCQ
Compression depth[g,h]	MCQ
Compression:ventilation ratio[g]	MCQ
Compression rate[g,h]	MCQ
Compression-only CPR[g,h,k]	Yes/No
Treatment of a choking patient, conscious, unable to either cough or talk[g]	MCQ
Self-assessed confidence in the ability to perform resuscitation	1-10 scale

[a]N/A: not applicable.
[b]Regex: regular expression validation.
[c]MCQ: multiple choice question (only one answer accepted).
[d]BLS: basic life support.
[e]ACLS: advanced cardiovascular life support.
[f]AED: automated external defibrillator.
[g]Questions used to compute the primary outcome (score out of 10 questions).
[h]Questions used to compute the "essential BLS knowledge" secondary outcome.
[i]MAQ: multiple answer question (more than one answer accepted).
[j]OHCA: out-of-hospital cardiac arrest.
[k]CPR: cardiopulmonary resuscitation.

Enrollment and Consent

As the UGFM BLS-AED course is based on a "flipped classroom" teaching format, medical students must complete an institutional, interactive, electronic learning (eLearning) module prior to attending their first 2-hour BLS-AED workshop. A link was therefore embedded in the very first slide of the eLearning module (Figure 1), prompting the 183 second-year UGFM students, which represented a convenience sample, to participate in the study before accessing the learning material.

The control group was composed of lay people attending a first aid course. These participants were recruited thanks to the Association Genevoise des Sections de Samaritains (AGSS), a

Red Cross–affiliated training center. The AGSS agreed to send a single mailing (Multimedia Appendix 2) on our behalf to each first aid course participant who registered between October 2019 and April 2020. In Switzerland, first aid courses are mandatory to obtain a driving license [18]. Though anonymity was ensured by virtue of the invitation method, no reminder could be sent as we did not have access to the participants' email addresses or to their identities. These invitation emails were sent at least 2 weeks before these courses were scheduled, and participants were specifically asked to take the questionnaire before attending the course.

Information regarding the study was displayed along with an electronic consent form before the questionnaire could be started [19]. Participants were informed of the study's purpose and of its estimated length. Identity and contact of the investigators were given, and information regarding data handling was provided. Participants were informed that no personal data would be asked for or recorded and that they would not be solicited any further after having completed the questionnaire.

No incentive was given to promote participation, which was not required to attend either course.

Figure 1. Electronic learning (eLearning) slide inviting medical students to participate in the study.

Data Collection

All questions had to be answered, and pages were to be entirely filled before participants could proceed to the following part of the survey. Answer consistency was checked using regular expression (regex) validation rules, and participants were warned whenever an inconsistent answer was identified. Participants could change their answers using a "back" button until the questionnaire was finalized.

Data were stored on an encrypted MySQL compatible database (MariaDB 5.5.5, MariaDB Foundation, Wakefield, MA) located on a Swiss server. As this was a closed survey, with the link only provided either on the first eLearning slide or in the emails sent by the AGSS, no specific tracking identifier (cookies or IP address) was used.

Outcomes

The primary outcome was the between-group difference on the 10-question quiz score. Each question was equally weighted and could only be considered as correct or incorrect. Thus, the total score was computed for each participant by summing all correct answers. Secondary outcomes were the differences in the rate of correct answers for each individual question and in the level of self-assessed confidence in the ability to perform resuscitation. We also computed a score dubbed "essential BLS knowledge," which is the sum of 6 critical questions related to BLS (Table 1) and follows the logic of the chain of survival. Indeed, the other questions used to compute the primary outcome are either related to other first aid maneuvers or are not deemed of critical importance to the adequate provision of resuscitative maneuvers.

Data Availability

The original data have been deposited to Mendeley Data [20].

Statistical Analysis

Statistical analysis was carried out using STATA 16.1 (StataCorp LLC, College Station, TX). Incomplete questionnaires, as well as those completed by BLS-AED instructors or already certified health care professionals, were excluded from the analysis.

The scores of the answers to the 10 predefined questions were added to compute the overall quiz score defined as the primary outcome (minimum = 0; maximum = 10). No differential weighting was applied, and each individual question was worth 1 point. The "essential BLS knowledge" score was computed in the exact same way (minimum = 0; maximum = 6).

Normality was assessed both graphically and by the Kolmogorov-Smirnov test. Analysis of continuous variables was first performed using the Student *t* test, and results were then adjusted according to age, gender, and prior BLS experience through the use of multivariable linear regression. Results are presented as mean (SD). Given the limited sample size, the Fisher exact test was used for between-groups comparison of binary variables. A double-sided *P* value <.05 was considered significant.

A subgroup analysis to identify a potential effect of having attended a BLS course prior to taking the survey was decided post hoc. A sensitivity analysis excluding medical students who had prior training as health care students or who were already certified rescuers was also performed.

Results

The participation rate was higher (*P*=.03) in the medical students' group (80/183, 44%) than in the control group (74/256, 29%). While no participant had to be excluded from the medical students' group, the exclusion criteria were met for 12 participants in the control group (Figure 2). Among the 80 medical students who completed the survey, 1 was a former nursing student, and 13 answered they were certified rescuers. There were no fully qualified health care professionals in this group.

Participants were older in the control group (mean 34.0 years, SD 12.7 years) than in the medical students' group (mean 22.5 years, SD 4.4 years). There was a majority of women in both groups (49/74, 66% in the control group and 58/80, 73% in the medical students' group) with no significant difference between groups (*P*=.35). While the proportion of participants who had already heard of BLS or of advanced cardiovascular life support was also similar (48/80, 60% of medical students vs 41/74, 55% of lay people; *P*=.39), medical students more often declared having already attended a BLS course (68/80, 85% vs 45/74, 61%, *P*<.001).

The mean score on the 10-question composite outcome was higher in medical students (mean 5.8, SD 1.7 vs mean 4.2, SD 1.7; *P*<.001), with 6 questions displaying significant differences (Table 2). Older participants were more likely to score lower, with a coefficient of –0.031 per year (95% CI –0.060 to –0.003, *P*=.03), while the participants who had attended a BLS course prior to taking the questionnaire scored higher (1.166, 95% CI 0.470 to 1.862, *P*=.001). No participant was able to correctly answer all 10 questions.

Medical students also scored better than lay people on the 6-question "essential BLS knowledge" score (mean 3.0, SD 1.1 vs mean 2.0, SD 1.0; *P*<.001). Older age was also correlated with lower scores (coefficient –0.025 per year, 95% CI –0.043 to –0.007, *P*=.007), and participants who had attended a BLS course prior to completing the questionnaire also scored higher on this index (0.698, 95% CI 0.258 to 1.138, *P*=.002). Once again, no participant was able to correctly answer all 6 questions. Neither score was affected by gender.

Lay people who had already attended a BLS course before participating in the survey did not perform better than those who had not (mean 4.5, SD 1.6 vs mean 3.8, SD 1.7, *P*=.14). The same held true for medical students (mean 5.9, SD 2.2 vs mean 5.4, SD 1.6, *P*=.39).

Medical students felt more confident than lay people in their ability to perform resuscitation (mean 4.7, SD 2.2 vs mean 3.1, SD 2.1, *P*<.001). Participants who had already attended a BLS course felt more confident than those who had not (coefficient 1.9, 95% CI 1.1 to 2.7, *P*<.001), while female participants felt less confident (coefficient –1.1, 95% CI –1.9 to –0.4, *P*=.003). Older participants felt less confident than their younger counterparts (coefficient –0.041 per year, 95% CI –0.074 to –0.009, *P*=.01). The proportion of participants wishing to receive more BLS training was higher in the group composed of medical students (70/80, 88% vs 42/74, 57%, *P*<.001).

Excluding medical students who were either former nursing students or certified rescuers neutralized the effect of age on the 10-question score and on the confidence but did not significantly change the magnitude or direction of the other results.

Figure 2. Study flowchart for (A) lay people and (B) medical students. AED: automated external defibrillator; BLS: basic life support.

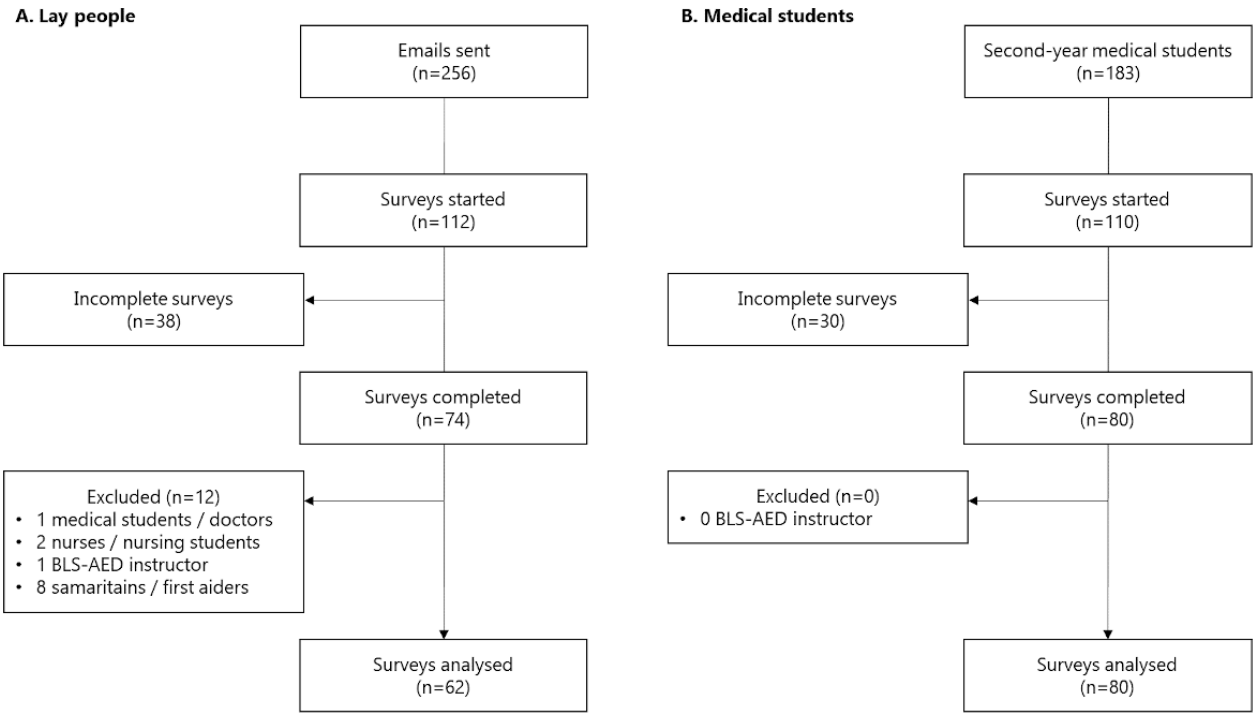


Table 2. Scores on the 10-question composite outcome by group.

Question	Lay people (n=62), n (%)	Medical students (n=80), n (%)	*P* value
Phone number of the emergency medical services dispatch central[a]	51 (82)	75 (94)	.06
Meaning of "AED"[b,c]	27 (44)	42 (53)	.31
Criteria used to recognize OHCA[d]	5 (8)	16 (20)	.06
BLS[e] sequence	2 (3)	0 (0)	.19
Artery for pulse assessment	50 (80)	50 (63)	.03
Compression depth	9 (14)	29 (36)	.004
Compression:ventilation ratio	21 (34)	62 (78)	<.001
Compression rate	15 (24)	41 (51)	.002
Compression-only CPR[f]	39 (63)	75 (94)	<.001
Treatment of a choking patient, conscious, unable to either cough or talk	43 (69)	75 (94)	<.001

[a]Accepted answers: 112, 144, and 911, which all work in Geneva, Switzerland; 144 is the official Swiss number.

[b]AED: automated external defibrillator.

[c]All answers containing "defibrillator" in English or in French were considered as correct (not case sensitive, spelling mistakes accepted).

[d]OHCA: out-of-hospital cardiac arrest.

[e]BLS: basic life support.

[f]CPR: cardiopulmonary resuscitation.

Discussion

Principal Findings

In this study, second-year medical students performed somewhat better than lay people, though neither group scored high on a simple 10-question quiz assessing BLS knowledge. The difference might arise from different grounds. First of all, medical students might indeed be more interested in this domain given their chosen profession. Moreover, though our regional policies have evolved little with regard to BLS training promotion, private initiatives have progressively surfaced in an attempt to increase general awareness regarding OHCA. As medical students were more than a decade younger than the

control group, they might have been more exposed to such campaigns and therefore more interested in this topic.

Our findings highlight weaknesses in the first 3 links of the chain of survival [21]. While the proportion of participants who correctly remembered the phone number of the medical emergency communication center was high in both groups, most were unable to correctly describe the criteria used to identify OHCA. Failure to correctly identify OHCA may lead to a delayed call to the dispatch center and thereby worsen the victim's prognosis [22]. Once called, dispatchers should nevertheless be able to help identify OHCA and lead the bystander to start appropriate actions [23]. However, while cardiopulmonary resuscitation (CPR) might be initiated, our survey results suggest that compression could be of limited quality since answers related to compression rate and depth were mostly incorrect. These findings are worrying since high-quality chest compressions are of paramount importance and have been shown to improve survival outcomes [24-26]. The third link in the chain, defibrillation, was only assessed with one simple question that focused on the meaning of the AED abbreviation. Less than half of all participants were able to determine that these 3 letters were commonly used to refer to a defibrillator. However, one could reasonably expect that associating the lightning pictograph to the abbreviation might improve identification of these devices [27]. Moreover, most emergency medical dispatchers are now trained to help bystanders or first responders localize, retrieve, and use AEDs [28].

Although lay rescuers, either trained or untrained, have not been expected to check for a pulse since the publication of the 2010 CPR guidelines [29], we nevertheless elected to ask a question regarding pulse assessment. Indeed, our BLS teaching faculty considers even junior medical students to be health care providers, and these students are therefore expected to know how to check for a pulse, particularly to assess a potential return of spontaneous circulation [30]. It must however be unequivocal that the pulse check should not be performed to diagnose cardiac arrest as it has been shown to be highly unreliable [31].

Even though emergency medical systems have evolved to overcome as much as possible the lack of BLS training [32], developing first aid knowledge and skills among the general population is essential to improve outcomes not only in OHCA victims but also in patients presenting with other acute illnesses or injuries [33]. Many different strategies can be considered to enhance the global level of awareness regarding OHCA and CPR, including systematic teaching of BLS maneuvers to school children [34-36]. This strategy has been shown to be particularly effective, as children are eager to transfer their acquired knowledge to their parents, siblings, and friends [37,38]. Training school children and providing them with cheap, basic manikins has been shown to help disseminate CPR training among their relatives [39], and many European countries have adhered to the "KIDS SAVE LIVES" statement, which aims to provide BLS education to schoolchildren aged 12 years and older [40]. Until such a change in BLS education policies occurs, medical students should be offered first aid courses earlier in their curriculum. The recently released European Resuscitation Council guidance note related to the CPR competencies required in undergraduate health care students could be used as a guideline to facilitate the implementation of BLS courses sooner in the undergraduate curriculum [41]. This note advocates, as we do, for the mandatory teaching of BLS maneuvers already during the first year of undergraduate education for students of health care professions.

However, months, if not years, might elapse before any change in public health policies can be accomplished, and modifying the undergraduate medical curriculum might take just as long. Therefore, to enhance the awareness of UGFM junior medical students regarding BLS issues, a team of senior medical students and faculty members developed an accelerated first aid course [42]. In January 2021, first-year medical students attended a brief intervention inviting them to complete an eLearning module. In the context of the current COVID-19 pandemic, asynchronous distance learning using interactive eLearning modules has been shown to enhance knowledge acquisition in medical students [43]. After completing this module, first-year medical students were invited to register for 1-hour practice sessions. Students who successfully complete this learning path will be able to enlist as first responders. Future studies should determine whether this strategy is successful and can improve BLS knowledge in this population.

The results of our post hoc analysis regarding the effect of having attended a BLS course prior to taking the survey are cause for concern. Indeed, participants who declared having attended such a course in the past did not perform significantly better than those who did not, regardless of the study group. While a change in the guidelines between a prior course and the moment of the survey could be hypothesized, the young age of the participants, particularly in the medical students' group, makes it unlikely. Low scores were indeed recorded regarding key elements that were already part of the 2010 guidelines (ie, the criteria used to recognize OHCA [29]). Many factors, such as the time elapsed since the last BLS course, the course format, or the number of previous courses the participants had attended, may explain the apparent lack of BLS knowledge retention [44-46]. However, as this was an unexpected finding, the design of our study was ill suited to explore its grounds, and further studies will be needed to confirm this result and try to determine potential causes.

This study has other limitations that should also be acknowledged. First, owing to the study design and to the impossibility of sending email reminders, the participation rate was rather low, particularly among lay rescuers. This might have led to an overestimation of BLS knowledge in both groups due to selection bias. The method of recruitment has been shown to significantly alter the participation rates [47,48], and previous studies have reported that paper questionnaires have slightly higher response rates than web-based ones [49]. The use of paper questionnaires is however associated with much higher costs and carries an increased risk of generating missing values.

In addition, we were unable to determine whether the questionnaire had actually been completed before the course. Nevertheless, medical students were required to complete the eLearning module with the embedded invitation slide before

attending their first BLS course, and lay people were sent the invitation email at least 2 weeks before attending their course.

Another limitation is that our control group cannot be considered as a true surrogate of the general population. Indeed, as participants were rather young and as BLS training initiatives have progressively increased, it is to be expected that BLS knowledge would actually be lower in a more representative sample of the general population. Another limitation is linked to the specificities of the UGFM curriculum. In our curriculum, medical students and dental medicine students share a common study path until the end of their second year of undergraduate training. Around one-fifth of second-year UGFM students are actually dental medicine students. There is however little reason to believe that their interest in BLS procedures should be different for these students compared to medical students whose interest is not in an acute medicine specialty. In addition, most junior medical students have not yet decided upon a specific career at this stage [50]. One last limitation is that no survey can be considered as a true assessment of the way bystanders would manage OHCA. Nevertheless, there is a correlation between confidence, CPR skills, and intention to perform resuscitation, and a low level of knowledge can hardly foster confidence [51].

Conclusion

Although medical students were more knowledgeable than lay people regarding BLS-AED procedures, their proportion of correct answers was still low. As OHCA recognition and high-quality chest compressions are paramount to increasing survival rates, a change in the curriculum, as well as a global transformation in the way the general population is educated regarding first aid maneuvers, could help improve outcomes.

Acknowledgments

The authors would like to thank the Association Genevoise des Sections de Samaritains (AGSS) and particularly Mrs Véronique Volken for having dispatched invitation emails on their behalf.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions and expected answers: original (French) version and English translation.
[PDF File (Adobe PDF File), 470 KB - jmir_v23i2e25125_app1.pdf]

Multimedia Appendix 2
Text sent in the invitation e-mails (dispatched on our behalf by the Association Genevoise des Sections de Samaritains).
[PDF File (Adobe PDF File), 50 KB - jmir_v23i2e25125_app2.pdf]

References

1. Kiyohara K, Sado J, Kitamura T, Ayusawa M, Nitta M, Iwami T, et al. Public-access automated external defibrillation and bystander-initiated cardiopulmonary resuscitation in schools: a nationwide investigation in Japan. Europace 2019 Mar 01;21(3):451-458. [doi: 10.1093/europace/euy261] [Medline: 30500911]
2. Rivera NT, Kumar SL, Bhandari RK, Kumar SD. Disparities in Survival with Bystander CPR following Cardiopulmonary Arrest Based on Neighborhood Characteristics. Emerg Med Int 2016;2016:6983750-6983758 [FREE Full text] [doi: 10.1155/2016/6983750] [Medline: 27379186]
3. Bray JE, Straney L, Smith K, Cartledge S, Case R, Bernard S, et al. Regions With Low Rates of Bystander Cardiopulmonary Resuscitation (CPR) Have Lower Rates of CPR Training in Victoria, Australia. JAHA 2017 Nov 06;6(6):1. [doi: 10.1161/jaha.117.005972] [Medline: 28584073]
4. Wietlisbach M, Schüpfer G. [Resuscitation]. Praxis (Bern 1994) 1997 Feb 04;86(6):182-188. [Medline: 9082524]
5. Larribau R, Deham H, Niquille M, Sarasin FP. Improvement of out-of-hospital cardiac arrest survival rate after implementation of the 2010 resuscitation guidelines. PLoS One 2018 Sep 24;13(9):e0204169 [FREE Full text] [doi: 10.1371/journal.pone.0204169] [Medline: 30248116]
6. Baldi E, Contri E, Bailoni A, Rendic K, Turcan V, Donchev N, et al. Final-year medical students' knowledge of cardiac arrest and CPR: We must do more!. Int J Cardiol 2019 Dec 01;296:76-80. [doi: 10.1016/j.ijcard.2019.07.016] [Medline: 31375334]
7. Roshana S, Kh B, Rm P, Mw S. Basic life support: knowledge and attitude of medical/paramedical professionals. World J Emerg Med 2012;3(2):141-145 [FREE Full text] [doi: 10.5847/wjem.j.issn.1920-8642.2012.02.011] [Medline: 25215053]
8. Kila T, Yockopua S. Knowledge of cardiopulmonary resuscitation among doctors at the Port Moresby General Hospital. P N G Med J 2012;55(1-4):76-87. [Medline: 25338477]
9. Almesned A, Almeman A, Alakhtar AM, AlAboudi AA, Alotaibi AZ, Al-Ghasham YA, et al. Basic life support knowledge of healthcare students and professionals in the Qassim University. Int J Health Sci (Qassim) 2014 Apr;8(2):141-150 [FREE Full text] [doi: 10.12816/0006080] [Medline: 25246881]

10. Willmore RD, Veljanoski D, Ozdes F, Stephens B, Mooney J, Crumley SG, et al. Do medical students studying in the United Kingdom have an adequate factual knowledge of basic life support? World J Emerg Med 2019;10(2):75-80 [FREE Full text] [doi: 10.5847/wjem.j.1920-8642.2019.02.002] [Medline: 30687442]
11. Xie JY, Frost R, Meakin R. Not quite a doctor, but should I help? A qualitative exploration of medical students' attitudes towards responding to medical emergencies that occur in the public domain. BMJ Open 2019 Apr 08;9(4):e028035 [FREE Full text] [doi: 10.1136/bmjopen-2018-028035] [Medline: 30962243]
12. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
13. SR 810.30 Federal Act of 30 September 2011 on Research involving Human Beings (Human Research Act, HRA). Swiss Confederation. 2011. URL: https://www.admin.ch/opc/en/classified-compilation/20061313/index.html [accessed 2020-05-12]
14. McCambridge J, Kalaitzaki E, White IR, Khadjesari Z, Murray E, Linke S, et al. Impact of length or relevance of questionnaires on attrition in online trials: randomized controlled trial. J Med Internet Res 2011 Nov 18;13(4):e96 [FREE Full text] [doi: 10.2196/jmir.1733] [Medline: 22100793]
15. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
16. Hochheimer CJ, Sabo RT, Perera RA, Mukhopadhyay N, Krist AH. Identifying Attrition Phases in Survey Data: Applicability and Assessment Study. J Med Internet Res 2019 Aug 23;21(8):e12811 [FREE Full text] [doi: 10.2196/12811] [Medline: 31444875]
17. Travers AH, Perkins GD, Berg R, Castren M, Considine J, Escalante R, Basic Life Support Chapter Collaborators. Part 3: Adult Basic Life Support and Automated External Defibrillation: 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations. Circulation 2015 Oct 20;132(16 Suppl 1):S51-S83. [doi: 10.1161/CIR.0000000000000272] [Medline: 26472859]
18. Regard S, Rosa D, Suppan M, Giangaspero C, Larribau R, Niquille M, et al. Evolution of Bystander Intention to Perform Resuscitation Since Last Training: Web-Based Survey. JMIR Form Res 2020 Nov 30;4(11):e24798 [FREE Full text] [doi: 10.2196/24798] [Medline: 33252342]
19. Eysenbach G, Wyatt J. Using the Internet for surveys and health research. J Med Internet Res 2002 Nov;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
20. Sturny L, Regard S, Schiffer E, Suppan L. Basic life support knowledge among junior medical students dataset. Mendeley Data 2020:1. [doi: 10.17632/99w7bb2p5v.1]
21. Perkins GD, Handley AJ, Koster RW, Castrén M, Smyth MA, Olasveengen T, Adult basic life support and automated external defibrillation section Collaborators. European Resuscitation Council Guidelines for Resuscitation 2015: Section 2. Adult basic life support and automated external defibrillation. Resuscitation 2015 Oct;95(10):81-99. [doi: 10.1016/j.resuscitation.2015.07.015] [Medline: 26477420]
22. Waalewijn RA, Tijssen JG, Koster RW. Bystander initiated actions in out-of-hospital cardiopulmonary resuscitation: results from the Amsterdam Resuscitation Study (ARRESUST). Resuscitation 2001 Sep;50(3):273-279. [doi: 10.1016/S0300-9572(01)00354-9] [Medline: 11719156]
23. Vaillancourt C, Verma A, Trickett J, Crete D, Beaudoin T, Nesbitt L, et al. Evaluating the effectiveness of dispatch-assisted cardiopulmonary resuscitation instructions. Acad Emerg Med 2007 Oct;14(10):877-883 [FREE Full text] [doi: 10.1197/j.aem.2007.06.021] [Medline: 17761545]
24. Stiell IG, Brown SP, Christenson J, Cheskes S, Nichol G, Powell J, Resuscitation Outcomes Consortium (ROC) Investigators. What is the role of chest compression depth during out-of-hospital cardiac arrest resuscitation? Crit Care Med 2012 Apr;40(4):1192-1198 [FREE Full text] [doi: 10.1097/CCM.0b013e31823bc8bb] [Medline: 22202708]
25. Vadeboncoeur T, Stolz U, Panchal A, Silver A, Venuti M, Tobin J, et al. Chest compression depth and survival in out-of-hospital cardiac arrest. Resuscitation 2014 Feb;85(2):182-188. [doi: 10.1016/j.resuscitation.2013.10.002] [Medline: 24125742]
26. Idris AH, Guffey D, Pepe PE, Brown SP, Brooks SC, Callaway CW, et al. Chest Compression Rates and Survival Following Out-of-Hospital Cardiac Arrest*. Critical Care Medicine 2015;43(4):840-848. [doi: 10.1097/ccm.0000000000000824] [Medline: 25565457]
27. Smith CM, Colquhoun MC, Samuels M, Hodson M, Mitchell S, O'Sullivan J. New signs to encourage the use of Automated External Defibrillators by the lay public. Resuscitation 2017 May;114:100-105. [doi: 10.1016/j.resuscitation.2017.03.012] [Medline: 28323083]
28. Stieglis R, Zijlstra JA, Riedijk F, Smeekes M, van der Worp WE, Koster RW. AED and text message responders density in residential areas for rapid response in out-of-hospital cardiac arrest. Resuscitation 2020 May;150:170-177 [FREE Full text] [doi: 10.1016/j.resuscitation.2020.01.031] [Medline: 32045663]
29. Koster RW, Baubin MA, Bossaert LL, Caballero A, Cassan P, Castrén M, et al. European Resuscitation Council Guidelines for Resuscitation 2010 Section 2. Adult basic life support and use of automated external defibrillators. Resuscitation 2010 Oct;81(10):1277-1292 [FREE Full text] [doi: 10.1016/j.resuscitation.2010.08.009] [Medline: 20956051]
30. Moule P. Checking the carotid pulse: diagnostic accuracy in students of the healthcare professions. Resuscitation 2000 May;44(3):195-201. [doi: 10.1016/s0300-9572(00)00139-8] [Medline: 10825620]

31. Ochoa F, Ramalle-Gómara E, Carpintero J, Garcı´a A, Saralegui I. Competence of health professionals to check the carotid pulse. Resuscitation 1998 Jun;37(3):173-175. [doi: 10.1016/s0300-9572(98)00055-0] [Medline: 9715777]
32. Tanaka Y, Taniguchi J, Wato Y, Yoshida Y, Inaba H. The continuous quality improvement project for telephone-assisted instruction of cardiopulmonary resuscitation increased the incidence of bystander CPR and improved the outcomes of out-of-hospital cardiac arrests. Resuscitation 2012 Oct;83(10):1235-1241. [doi: 10.1016/j.resuscitation.2012.02.013] [Medline: 22366353]
33. Zideman DA, De Buck ED, Singletary EM, Cassan P, Chalkias AF, Evans TR, et al. European Resuscitation Council Guidelines for Resuscitation 2015 Section 9. First aid. Resuscitation 2015 Oct;95:278-287. [doi: 10.1016/j.resuscitation.2015.07.031] [Medline: 26477417]
34. Baznik S, Mohar A. Basic life support training for primary school children. Resuscitation 2019 Sep;142:e55. [doi: 10.1016/j.resuscitation.2019.06.132]
35. Cave DM, Aufderheide TP, Beeson J, Ellison A, Gregory A, Hazinski MF, American Heart Association Emergency Cardiovascular Care Committee, Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation, Council on Cardiovascular Diseases in the Young, Council on Cardiovascular Nursing, Council on Clinical Cardiology, Advocacy Coordinating Committee. Importance and implementation of training in cardiopulmonary resuscitation and automated external defibrillation in schools: a science advisory from the American Heart Association. Circulation 2011 Feb 15;123(6):691-706. [doi: 10.1161/CIR.0b013e31820b5328] [Medline: 21220728]
36. Böttiger BW, Van Aken H. Kids save lives--Training school children in cardiopulmonary resuscitation worldwide is now endorsed by the World Health Organization (WHO). Resuscitation 2015 Sep;94:A5-A7. [doi: 10.1016/j.resuscitation.2015.07.005] [Medline: 26209417]
37. Dillon JJ. The role of the child in adult development Internet. Journal of Adult Development 2002;9:267-275. [doi: 10.1023/A:1020286910678]
38. Bresee S, Caruso B, Sales J, Lupele J, Freeman M. 'A child is also a teacher': exploring the potential for children as change agents in the context of a school-based WASH intervention in rural Eastern Zambia. Health Educ Res 2016 Aug 19;31(4):521-534. [doi: 10.1093/her/cyw022] [Medline: 27206442]
39. Isbye DL, Rasmussen LS, Ringsted C, Lippert FK. Disseminating cardiopulmonary resuscitation training by distributing 35,000 personal manikins among school children. Circulation 2007 Sep 18;116(12):1380-1385. [doi: 10.1161/CIRCULATIONAHA.107.710616] [Medline: 17724257]
40. Böttiger BW, Semeraro F, Altemeyer K, Breckwoldt J, Kreimeier U, Rücker G, et al. KIDS SAVE LIVES: School children education in resuscitation for Europe and the world. Eur J Anaesthesiol 2017 Dec;34(12):792-796. [doi: 10.1097/EJA.0000000000000713] [Medline: 29087993]
41. Baldi E, Savastano S, Contri E, Lockey A, Conaghan P, Hulme J, et al. Mandatory cardiopulmonary resuscitation competencies for undergraduate healthcare students in Europe: A European Resuscitation Council guidance note. Eur J Anaesthesiol 2020 Oct;37(10):839-841. [doi: 10.1097/EJA.0000000000001272] [Medline: 32925434]
42. Suppan L, Herren T, Taramarcaz V, Regard S, Martin-Achard S, Zamberg I, et al. A Short Intervention Followed by an Interactive E-Learning Module to Motivate Medical Students to Enlist as First Responders: Protocol for a Prospective Implementation Study. JMIR Res Protoc 2020 Nov 06;9(11):e24664 [FREE Full text] [doi: 10.2196/24664] [Medline: 33155574]
43. Suppan M, Stuby L, Carrera E, Cottet P, Koka A, Assal F, et al. Asynchronous Distance Learning of the National Institutes of Health Stroke Scale During the COVID-19 Pandemic (E-Learning vs Video): Randomized Controlled Trial. J Med Internet Res 2021 Jan 15;23(1):e23594 [FREE Full text] [doi: 10.2196/23594] [Medline: 33428581]
44. Nishiyama C, Iwami T, Murakami Y, Kitamura T, Okamoto Y, Marukawa S, et al. Effectiveness of simplified 15-min refresher BLS training program: a randomized controlled trial. Resuscitation 2015 May;90:56-60. [doi: 10.1016/j.resuscitation.2015.02.015] [Medline: 25724354]
45. Niles D, Sutton R, Donoghue A, Kalsi M, Roberts K, Boyle L, et al. "Rolling Refreshers": a novel approach to maintain CPR psychomotor skill competence. Resuscitation 2009 Aug;80(8):909-912. [doi: 10.1016/j.resuscitation.2009.04.021] [Medline: 19467759]
46. Kovács E, Jenei ZM, Csordás K, Fritúz G, Hauser B, Gyarmathy VA, et al. The timing of testing influences skill retention after basic life support training: a prospective quasi-experimental study. BMC Med Educ 2019 Dec 04;19(1):452 [FREE Full text] [doi: 10.1186/s12909-019-1881-7] [Medline: 31801502]
47. Loxton D, Harris ML, Forder P, Powers J, Townsend N, Byles J, et al. Factors Influencing Web-Based Survey Response for a Longitudinal Cohort of Young Women Born Between 1989 and 1995. J Med Internet Res 2019 Mar 25;21(3):e11286 [FREE Full text] [doi: 10.2196/11286] [Medline: 30907739]
48. Christensen T, Riis AH, Hatch EE, Wise LA, Nielsen MG, Rothman KJ, et al. Costs and Efficiency of Online and Offline Recruitment Methods: A Web-Based Cohort Study. J Med Internet Res 2017 Mar 01;19(3):e58 [FREE Full text] [doi: 10.2196/jmir.6716] [Medline: 28249833]
49. Ebert JF, Huibers L, Christensen B, Christensen MB. Paper- or Web-Based Questionnaire Invitations as a Method for Data Collection: Cross-Sectional Comparative Study of Differences in Response Rate, Completeness of Data, and Financial Cost. J Med Internet Res 2018 Jan 23;20(1):e24 [FREE Full text] [doi: 10.2196/jmir.8353] [Medline: 29362206]

50. Ock M, Han Y, Choi EY, Pyo J, Lee W. Perceptions of Medical Students Regarding Career Counseling in Korea: A Qualitative Study. Int J Environ Res Public Health 2020 May 16;17(10):3486 [FREE Full text] [doi: 10.3390/ijerph17103486] [Medline: 32429488]
51. Vaillancourt C, Stiell IG, Wells GA. Understanding and improving low bystander CPR rates: a systematic review of the literature. CJEM 2008 Jan;10(1):51-65. [doi: 10.1017/s1481803500010010] [Medline: 18226319]

Abbreviations

AED: automated external defibrillator
AGSS: Association Genevoise des Sections de Samaritains
BLS: basic life support
CPR: cardiopulmonary resuscitation
eLearning: electronic learning
OHCA: out-of-hospital cardiac arrest
UGFM: University of Geneva Faculty of Medicine

Edited by R Kukafka; submitted 19.10.20; peer-reviewed by E Baldi, C Beaudoin, D Millenaar; comments to author 23.12.20; revised version received 06.01.21; accepted 31.01.21; published 23.02.21.

Please cite as:
Sturny L, Regard S, Larribau R, Niquille M, Savoldelli GL, Sarasin F, Schiffer E, Suppan L
Differences in Basic Life Support Knowledge Between Junior Medical Students and Lay People: Web-Based Questionnaire Study
J Med Internet Res 2021;23(2):e25125
URL: https://www.jmir.org/2021/2/e25125
doi:10.2196/25125
PMID:33620322

Original Paper

Development and Feasibility of a Web-Based Decision Aid for Patients With Ulcerative Colitis: Qualitative Pilot Study

Andrew H Kim[1,2], MD; Afaf Girgis[1], PhD; Peter De Cruz[3,4], PhD, MD; Corey A Siegel[5], MD; Neda Karimi[1], PhD; Sasha O Ruban[1], BA; Alexandra J Sechi[2], RN; Wa Sang Watson Ng[1,2], PhD, MD; Jane M Andrews[6,7], PhD, MD; Susan J Connor[1,2], PhD, MD

[1]Ingham Institute for Applied Medical Research, South Western Sydney Clinical School, The University of New South Wales, Sydney, Australia
[2]Department of Gastroenterology, Liverpool Hospital, Sydney, Australia
[3]Department of Gastroenterology, The Austin Hospital, Melbourne, Australia
[4]Department of Medicine, Austin Academic Centre, University of Melbourne, Melbourne, Australia
[5]Section of Gastroenterology and Hepatology, Dartmouth-Hitchcock Medical Center, Lebanon, NH, United States
[6]IBD Service, Department of Gastroenterology & Hepatology, Royal Adelaide Hospital, Adelaide, Australia
[7]Faculty of Medicine, University of Adelaide, Adelaide, Australia

Corresponding Author:
Susan J Connor, PhD, MD
Ingham Institute for Applied Medical Research
South Western Sydney Clinical School
The University of New South Wales
Department of Gastroenterology Liverpool Hospital
Sydney
Australia
Phone: 61 2 8738 4085
Email: Susan.Connor1@health.nsw.gov.au

Abstract

Background: Shared decision making (SDM) is becoming an important part of ulcerative colitis (UC) management because of the increasing complexity of available treatment choices and their trade-offs. The use of decision aids (DA) may be effective in increasing patients' participation in UC management but their uptake has been limited due to high attrition rates and lack of a participatory approach to their design and implementation.

Objective: The primary aim of this study is to explore the perspectives of Australian patients and their clinicians regarding the feasibility and acceptability of myAID, a web-based DA, in informing treatment decisions in UC. The secondary aim is to use the findings of this pilot study to inform the design of a cluster randomized clinical trial (CRCT) to assess the efficacy of the DA compared with usual care.

Methods: myAID, a DA was designed and developed using a participatory approach by a multidisciplinary team of clinicians, patients, and nonmedical volunteers. A qualitative pilot study to evaluate the DA, involving patients with UC facing new treatment decisions and inflammatory bowel disease clinicians, was undertaken.

Results: A total of 11 patients with UC and 15 clinicians provided feedback on myAID. Themes explored included the following: Acceptability and usability of myAID—myAID was found to be acceptable by the majority of clinicians as a tool to facilitate SDM, uptake was thought to vary depending on clinicians' approaches to patient education and practice, potential to overcome time restrictions associated with outpatient clinics was identified, presentation of unbiased information enabling patients to digest information at their own pace was noted, and potential to provoke anxiety among patients with a new diagnosis or mild disease was raised; Perceived role and usefulness of myAID—discordance was observed between patients who prioritized voicing preferences and clinicians who prioritized treatment adherence, and myAID facilitated early discussion of medical versus surgical treatment options; Target population and timing of use—greatest benefit was perceived at the time of initiating or changing treatment and following commencement of immunosuppressive therapy; and Potential concerns and areas for improvement—some perceived that use of myAID may precipitate anxiety by increasing decisional conflict and impact the therapeutic relationship between patient and the clinician and may increase resource requirements.

Conclusions: These preliminary findings suggest that patients and clinicians consider myAID as a feasible and acceptable tool to facilitate SDM for UC management. These pilot data have informed a participatory approach to the design of a CRCT, which will evaluate the clinical efficacy of myAID compared with usual care.

Trial Registration: Australian New Zealand Clinical Trial Registry ACTRN12617001246370; http://anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=12617001246370.

(J Med Internet Res 2021;23(2):e15946) doi:10.2196/15946

KEYWORDS

shared decision making; decision aid; ulcerative colitis

Introduction

Background

Ulcerative colitis (UC) is an idiopathic chronic inflammatory condition involving the gastrointestinal tract, and it is characterized by relapsing and remitting symptoms of intestinal inflammation. Patients require long-term medical treatment, and in severe cases, they may require surgery to remove the entire colon. The trade-offs involving the available treatment options and their risk versus benefits are complex. Patients' values and preferences heavily influence treatment choice and adherence [1]. Therefore, decision-making regarding therapeutics in UC encompasses not only the identification of the best treatment strategy for individual patients to maximize important outcomes such as quality of life and better disease control but also patient education, patient engagement, and effective communication to help promote adherence [2].

Shared decision making (SDM) is considered an important component of patient-centered care that enables and encourages patients to participate actively in the management of their health [3]. Such an approach has been found to result in better health outcomes and health care experience as well as increased treatment adherence [4]. The goal of SDM is that clinicians and patients will share their knowledge, values, and preferences and deliberate together on management decisions. Moreover, a recent survey suggests that most patients with UC wish to participate in SDM with their gastroenterologist when making treatment decisions [5].

Although SDM has a number of purported benefits, it requires a considerable amount of time and effort from the clinician to discuss the treatment options and explain their benefits and risks, while also helping the patient to identify which option best matches their preferences. With the increasing burden and incidence of UC, SDM is becoming an increasingly challenging task in a time-pressured, resource-limited clinic environment. Therefore, more effective ways of communication and patient engagement are needed.

Decision aids (DAs) are useful decision support interventions that may help to overcome the barriers to SDM by providing patients with easy-to-understand, evidence-based information about their available options outside the consultation timeframe, encouraging active engagement in the decision-making process and prompting patients to think through their values and preferences and consider them in making a decision [6]. The use of DAs in other chronic diseases has been associated with increased patient knowledge, less decisional conflict, and fewer patients remaining undecided or passive in the decision-making process [7]. DAs are available in various formats, ranging from simple pamphlets to elaborate videos, with web-based DAs becoming increasingly popular due to low production cost, ease of updates, and patient reliance on the internet for information [5,8].

Despite their potential to aid in participatory medicine, the application of DAs in UC management to date has been limited to decisions regarding surgical treatment [9,10], and available eHealth technologies incorporating self-monitoring and self-management functionalities have experienced high attrition rates, preventing their widespread uptake in clinical practice [11]. Importantly, many of these interventions have lacked a participatory health research design to maximize the participation of patients and clinicians in their design and development [12,13]. Furthermore, patient and clinician perspectives on the best approach to the use of these tools in routine clinical practice are currently poorly understood.

Research Objectives

The primary objective of this study is to explore the perspectives of Australian patients and their clinicians regarding a web-based DA in informing treatment decisions in UC and investigate how such a tool may be best incorporated and used in clinical practice. The secondary objective is to use the findings of this pilot study to inform the design of a cluster randomized clinical trial (CRCT) aiming to assess the efficacy of myAID compared with usual care.

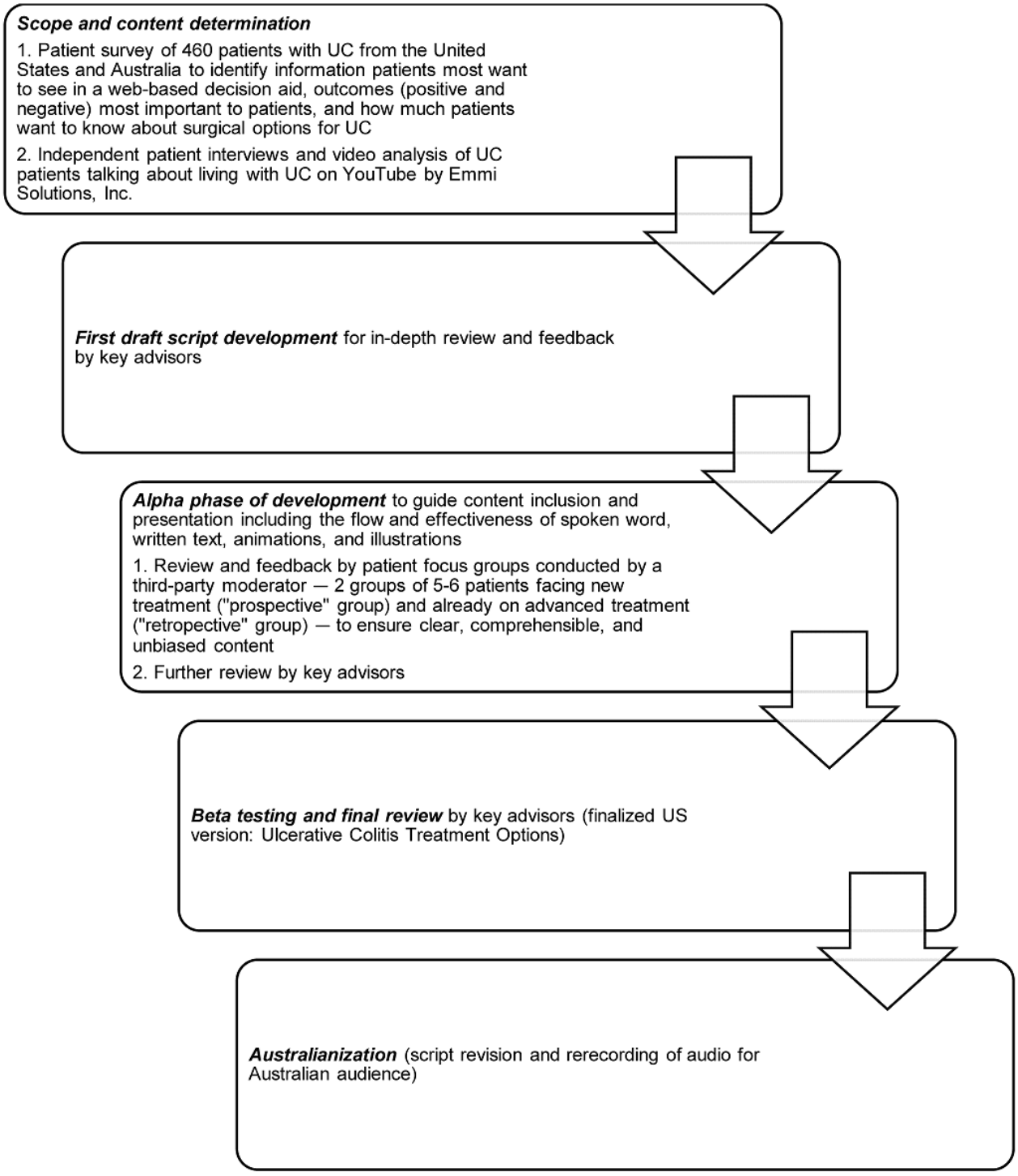
Methods

Development and Key Features of myAID

myAID is a web-based multimedia DA. It is the Australianized version of the original US program named Ulcerative Colitis Treatment Options (Emmi Solutions, Chicago) designed to facilitate and support SDM for treatment decisions in UC management. Throughout its development and design, a participatory approach [12,13] was undertaken with direct input from patients and a multidisciplinary panel of clinical experts (inflammatory bowel disease [IBD] physicians, colorectal surgeons, and experts on SDM experienced with DA use; see *Acknowledgments* section). This process involved close reference to the International Patient Decision Aids Standards checklist [14] and a rigorous evaluation process including in-house and independent reviews (Figure 1) [15]. The web-based format was selected based on a previous survey involving patients with UC from both the United States and

Australia [5]. The DA script and audio were *Australianized* (using Australian language and correct Australian medication names) through feedback from 1 patient, 4 nonmedical volunteers, 1 psychologist (AG), 3 Australian IBD physicians (SC, WN, and JA), and 2 Australian IBD nurses (As and ES), with rerecording of audio with the Australian script by Emmi Solutions in collaboration with Medibank Private Australia.

Figure 1. Development of the web-based decision aid.



myAID is web accessible (but not downloadable as a portable app) via a computer or a laptop with an internet connection, is available in English, and takes approximately 32 min if viewed uninterrupted. The content is organized into chapters (Textbox 1), which can be viewed as often as desired, and starts with information including common symptoms, treatment goals, and benefits and risks associated with well-controlled versus poorly controlled disease. It then presents a succinct summary of currently available medical and surgical treatments and their potential benefits and risks and a disclaimer confirming no support or influence from pharmaceutical companies. myAID has several interactive components that prompt patients to consider their personal treatment goals by asking questions about their current symptoms and concerns. They can select the

specific details or information they wish to view about a particular aspect of treatment, such as information regarding surgery, including the details of the steps involved in creating a J-pouch or stoma and the potential complications such as pouchitis or peristomal hernia. Patients can take virtual notes during the video, print these along with a summary of the information presented in myAID, and return and skip to specific chapters after viewing the entire video. Refer to Figure 2 for sample myAID screenshots.

Textbox 1. Summary of myAID chapters.

Ulcerative colitis

- An overview of ulcerative colitis, including anatomy, disease pathogenesis, typical symptoms, and treatment goals

Medical treatments

- An overview of available treatment options including brief discussions on the role of natural therapy including diet and probiotics as well as smoking

Considering medications

- Describes the importance of adherence and risk of flare when treatment is suddenly stopped, costs, and pros and cons of each medical (nonsurgical) treatment option including 5-aminosalicylates, steroids, thiopurines, and biologics (infliximab, adalimumab, golimumab, and vedolizumab)

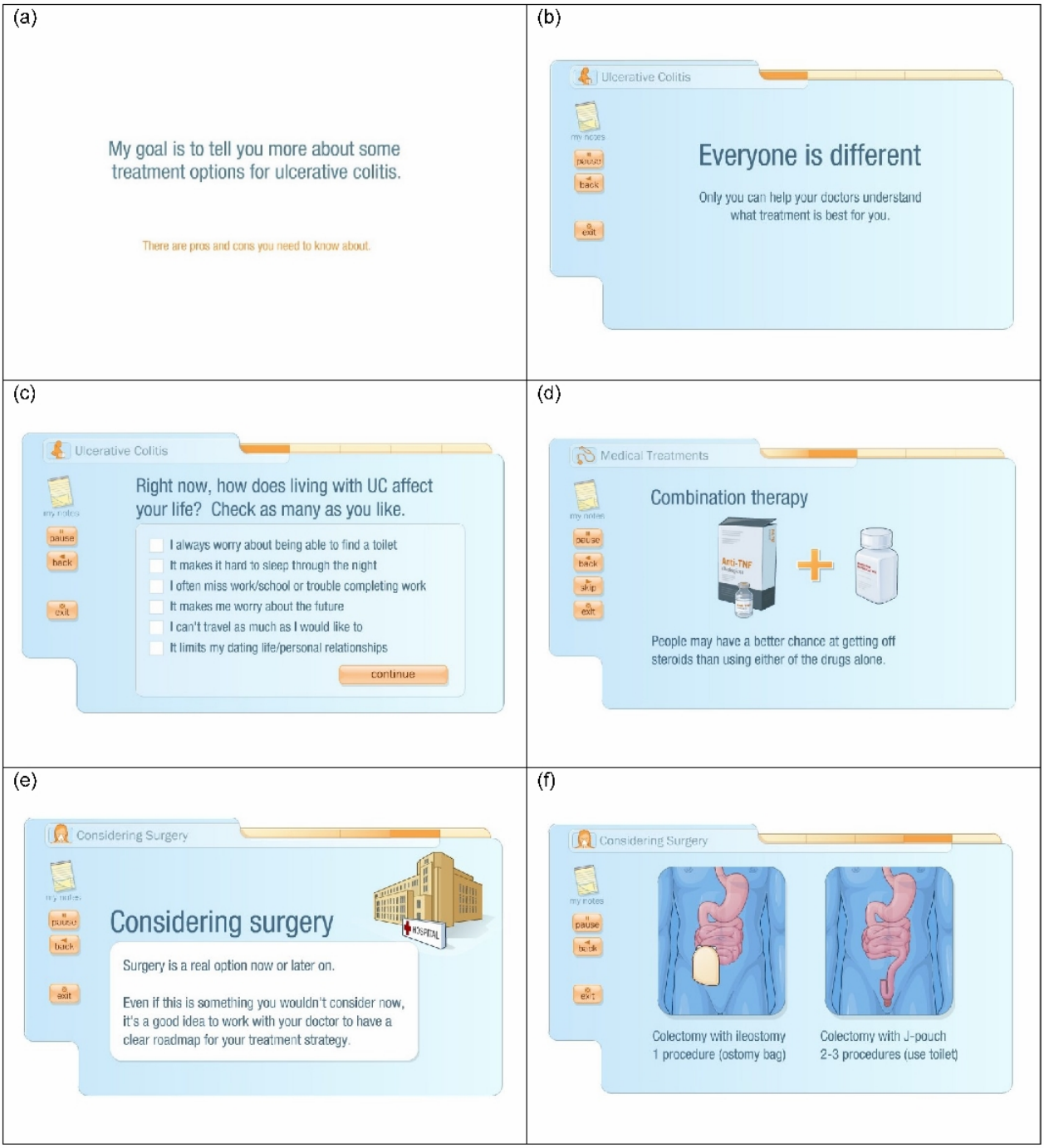
Considering surgery

- Describes surgery as a treatment option including the reasons to consider surgery, benefits, what to expect after surgery, and the potential risks and also discusses ileostomy versus J-pouch formation

Thinking it through

- A review of treatment options covered in previous sections including how they are given, how quickly they work, efficacy, risk of infection and lymphoma, and ileostomy versus J-pouch

Figure 2. Sample screenshots of myAID.



Study Participants and Recruitment

Patients

Patients aged above 18 years with UC who needed to make a new treatment decision after a previous or current trial of 5-aminosalicylate (5-ASA) therapy were eligible to participate. Eligible patients were consecutively identified and referred to the research assistant (RA) by their clinicians during their outpatient clinic attendance at Liverpool Hospital, New South Wales (NSW). The RA provided patients with a participant information sheet, explained study participation, and answered any questions before obtaining written consent.

Clinicians

As this was a feasibility study informing the CRCT, only clinicians from sites that had been allocated to the CRCT intervention arm and had access to myAID were eligible to participate. Of the 25 eligible clinicians, 15 were randomly selected, and all the clinicians consented.

Study Design

All consenting patients completed surveys to provide relevant sociodemographic and clinical information. Disease activity was determined by the Simple Clinical Colitis Activity Index (SCCAI) [16], where the score ranges from 0 to 19, with higher scores indicating greater disease severity and remission defined

as a score ≤2. Patients viewed myAID in the clinic (using an on-site computer) or at home (on their personal computer) via a URL unique to each patient. Patients returned to their clinician within 2 weeks for a follow-up consultation to discuss and decide their UC management, after which they participated in a one-to-one structured telephone interview (Multimedia Appendix 1) with the RA to provide feedback about myAID. Interviews (approximately 20 min) were audio-recorded and transcribed verbatim for qualitative analysis.

Participating clinicians were provided with the myAID URL and contacted by the RA within 2 weeks to confirm that they had viewed myAID. A 20-min structured telephone interview was then scheduled with the RA or lead researcher (AK) to provide feedback about myAID; interviews were audio-recorded and transcribed verbatim for qualitative analysis.

Successive transcribed interviews were reviewed, and their content was analyzed independently by 2 researchers (AK and SR) using thematic analysis [17]. Both deductive and inductive approaches were used to identify, report, and categorize themes or patterns and to ensure saturation is achieved. Subsequently, the themes were further refined and reviewed to ensure reviewers' agreement. Exemplary quotes were identified to denote each emerging theme during the analysis.

This study and the CRCT were approved by the Human Research Ethics Committee of South Western Sydney Local Health District and registered with the Australian New Zealand Clinical Trial Registry—ACTRN12617001246370.

Results

Participant Characteristics

A total of 11 patients were approached, and all of them were subsequently recruited (November 2015 to February 2016). The sample size was determined using the concept of thematic saturation. The majority of patients had left-sided disease (involvement limited to the proportion of the colon distal to the splenic flexure) for 2-10 years with a mean SCCAI of 6.5 (SD 5.6). A total of 15 clinicians with subspecialty interest in IBD participated (July to September 2016), the majority with ≥10 years of clinical experience and seeing ≥26 IBD patients per month. Tables 1 and 2 summarize the participant characteristics.

Table 1. Patient characteristics (N=11).

Characteristics	Total, n (%)
Age group (years)	
18-29	3 (27)
30-39	4 (36)
40-49	1 (9)
50-59	3 (27)
Gender	
Male	6 (55)
Female	5 (45)
Ethnicity	
White	5 (45)
Other	6 (55)
Education status	
Did not complete high school	2 (18)
High school or equivalent	5 (45)
Bachelor's degree or higher	4 (36)
Employment status	
Full-time	4 (36)
Part-time	2 (18)
Self-employed	2 (18)
Homemaker	1 (9)
Did not answer	1 (9)
Disabled or unable to work	1 (9)
Years since UC[a] diagnosis	
<2 years	1 (9)
2-10 years	10 (91)
Disease extent	
Rectum only	1 (9)
Left-sided	8 (73)
Most or all of colon affected	2 (18)
Previous treatment	
5-ASAs[b]	11 (100)
Prednisone	7 (64)
Thiopurines	3 (27)
Infliximab	2 (18)

[a]UC: ulcerative colitis.

[b]5-ASA: 5-aminosalicylates.

Table 2. Clinician characteristics (N=15).

Characteristics	Total, n (%)
Age (years)	
30-39	6 (40)
40-49	4 (27)
50-59	5 (33)
Gender	
Male	11 (73)
Female	4 (27)
Clinical setting	
Public only	1 (7)
Private only	0 (0)
Both	14 (93)
Access to IBD[a] nurse	
Full-time equivalent	11 (73)
Part time	4 (27)
None	0 (0)
Clinical experience (years)	
<10	6 (40)
10-20	5 (33)
>20	4 (27)
IBD patients seen per month	
≤25	1 (7)
26-75	8 (53)
≥76	6 (40)

[a]IBD: inflammatory bowel disease.

Patient Interviews

Of the 11 patients, 10 provided feedback about the acceptability and usability of myAID, perceived role and usefulness of myAID, target population and timing and place of use, and potential concerns and areas for improvement (refer to Table 3 for subthemes and exemplar quotes).

Table 3. Patient feedback on acceptability of myAID.

Feedback[a]	Quote
Acceptability and usability of myAID	
Easy to access, navigate, and use	• "Everything was easy." • "I could go into more detail if I wanted to."
Information presented is adequate, well delivered, and unbiased	• "It was spot on. It was perfect. I did not feel overwhelmed by it. But there was enough for me to take away and want to know more."
Content is informative and useful for patients	• "Half the questions I was going to ask her were answered in that video. Helped me understand my situation and what the doctor was saying too."
Perceived role and usefulness of myAID	
Education (especially in terms of surgical treatment and combination therapy) and promoting adherence	• "I had heard about things before and now I know what it is." • "now I am more aware of my options." • "I didn't know about injections or surgery [before]." • "I never took [the tablets] regularly. Now I will take them as doctor has prescribed."
Communication of concerns and participation in decision-making	• "It did help me discuss my concerns and options with her [my doctor]." • "Now I know what to talk about and what the doctor is saying and also my concerns."
Decision support	• "It confirmed what I wanted to do and confirmed my decision. It was a big help."
Psychological support (reducing anxiety)	• "I was scared to try other things before." • "I am more confident about my treatment."
Sharing information with family or other	• "It can be explained to them in a way I wouldn't be able to."
Suggested target population and timing or place of use	
Applicable at different stages of the treatment journey but may require caution in those newly diagnosed	• "Overall the whole video was helpful it told you everything, all worst case scenario and everything [so] you know what is happening" • "New people just diagnosed, [it] could scare them."
Suggested improvements and potential concerns	
Potential to increase anxiety and decisional conflict or burden—need for support from clinicians	• "Like surgery and stuff it could be scary." • "None of them is what I want. When you only have so many options you have to take one and that's the one I prefer more than the rest of them." • "...can be a bit of a shock too so maybe more help or support."
External factors may influence decisions—limitation of health care structure	• "To show government we tried this option before we go to other option [of my choice]." (on Pharmaceutical Benefits Scheme restrictions)
Suggestion for additional content	• "It would be so good to know what to eat–do a video about that"
Improved accessibility (on mobile devices)	• "I wanted to do it on my mobile but I had to do it on my computer at home."

[a]Themes arising from qualitative analysis.

Acceptability and Usability

Of the 11 patients, 8 viewed myAID in the clinic and 3 at home. Of the 10 interviewed patients, 9 watched myAID in one sitting, and all 10 patients found it easy to access myAID and navigate its contents and interactive components on the web-based platform; reported the amount of information presented as adequate and comprehensive; and found the information to be fairly presented (without bias), clear, and easy to understand. The majority of patients reported that myAID helped them with new knowledge about their disease and treatment options, leading to a much better understanding of the recent discussion with their clinicians. Many of these patients reported that their ability to contribute to the discussion was limited previously.

The sense of control over the viewed content was received positively, and patients reported that this feature allowed them to digest potentially confronting information at their own pace.

Perceived Role and Usefulness

All patients considered myAID as a useful educational tool that helped to improve knowledge and understanding of UC and available treatment options as well as the rationale for maintenance treatment and adherence. myAID was perceived to improve their ability to communicate their concerns and participate in decision-making more effectively and confidently. Viewing myAID was believed to help most patients by reaffirming their treatment decision or helping them change it with increased confidence. Sections reported as particularly helpful were the discussion on surgical treatment and information regarding combination therapy (using an immunomodulator together with a biologic drug). The information presented and the sense of control it conveyed were reported to improve patients' confidence and reduce their anxiety about new treatments. The majority of patients reported that they would share myAID with their family and relatives, explaining that they now had a way to explain their condition to a layperson in a language they can understand.

Target Population and Timing and Place of Use

Although myAID was considered acceptable and equally helpful for patients in various stages of UC in terms of disease duration and activity, some patients suggested caution against use in newly diagnosed patients for concerns of increasing anxiety, specifically referring to information about surgery. Several patients suggested that having their own time to digest the information away from the clinic was helpful. myAID was considered useful to view before consulting their clinician to understand their current situation and available treatment options and also afterward to help reaffirm decisions already made when introducing new treatment.

Potential Concerns and Areas for Improvement

Although patients indicated that myAID delivered new knowledge about their disease and available treatment options, they felt that information alone did not remove decisional conflict or burden and that they would like to follow this up with their clinician for support in the decision-making process. Inclusion of information about diet and enabling access to mobile devices were also desired by patients, the latter being thought of as a potentially more effective way of sharing information with others. One concern raised was that, in Australia, the Pharmaceutical Benefits Scheme influenced and limited treatment choices irrespective of patients' and clinicians' decisions (trial of thiopurines or methotrexate is mandated before funded biologic therapy for UC in Australia).

Clinician Interviews

Subthemes and exemplar clinician quotes are provided in Table 4.

Table 4. Clinician feedback on acceptability of myAID.

Feedback[a]	Quote
Acceptability and usability of myAID from a clinician perspective	
Information is tailored for individual patients	• "I can see benefit in people being able to go through this more than once, they could pick and choose what to view and not have to see what is not relevant anymore." • "You could hear it or not. I thought that was quite well done."
Information presented is adequate and delivered in a language that is suited to patients	• "It was quite thorough without being over the top." • "Overall I really liked the non-confrontational way it described things and the way that it sort of went through educating people in language that is so accessible and non-threatening."
Content is useful and relevant for patients	• "It covered a lot of very frequent questions that we get." • "The impression that I get is that it also draws on quite a lot of info from patients and their perspective."
Perceived role and usefulness of myAID from a clinician perspective	
Promoting treatment adherence	• "...video is quite good at highlighting some of those points like 'if you stop your medication it comes back'."
Patient education	• "I thought it was done in a language that's very easy for a patient to understand" • "...they need time to digest information and then think about how they are going to use it"
Potential to improve patient engagement and participation in decision-making	• "We spend a lot of time talking in clinic and it would be nice to back it up and they can go through it in their own time." • "Sometimes it's like we are putting them on the spot to make a decision (on the traditional approach to decision-making)"
Suggested target population and timing or place of use	
Applicable at different stages of the treatment journey but particularly at the time of treatment escalation	• "Depends on the course you are–at new diagnosis or 5 years down the track. Decision-making and info requirements are substantially different." • "Definitely early in their diagnosis and then the way it is structured it would be great at time of change of treatment or escalation." • "If someone presented with very severe disease you might want to show them that upfront because that would be incredibly helpful when you are scared and you don't know what's going on and you can see what options you have."
Early introduction may be beneficial but will require caution in those newly diagnosed	• "I think so much depends on the level of education receives at the beginning and certainly time of diagnosis is a critical time." • "I don't think at diagnosis because they are often overwhelmed with a lot of information." • "…people are often quite overwhelmed after diagnosis. So I would use it in first remission...and then each time there is a change or up titration."
Preferred setting for use is at home to allow for time and discussion but it could also be used in clinic	• Home: "They need time to digest information and then think about how they are going to use it." • "I think the home is the best place so they can take their time and involve other people if they want." • Clinic: "One option to start it in clinic while they are waiting for us."
Support is needed to address questions promptly	• "I think that people who have questions need to have them answered promptly." • "IBD nurse needs to be available otherwise questions will be forgotten about."
Suggested improvements and potential concerns	

Feedback[a]	Quote
Potential for information overload and intimidation—need for careful selection of patients	• "It would be quite an intimidating thing for new patients to be confronted with things like surgery." • "Not having to hear about detail about surgery that they might not want to hear from a computer for the first time."
Potential for decisional conflict and disagreement	• "Patients feeling like they have to decide. If they come to me and say I want this and if I don't think that's right it could create a problem."
Potential for outdated information	• "Challenges are always to keep info up to date as we get new drug approvals."
Potential for increased resource use	• "Making sure that there is a short time between them viewing it and coming back...if you leave it too long then the opportunity is lost...I don't think it should be more than a couple of weeks."
Suggestion for additional content, features, and structural elements; reference to patient support programs; improved navigation and ability to print; and improved access	• "Challenges are always to keep info up to date as we get new drug approvals." • "...have a scoring system so when they come back to clinic after seeing it, they have a validated index to then discuss. Could be clinical activity index or a PRO (patient reported outcome)." • "Link to patient support programs." • "they could print something out at the end so they could discuss it with their doctors." • "Patients are always on their mobiles. It's easier to have time. Needs to be mobile able."

[a]Themes arising from qualitative analysis.

Acceptability and Usability

myAID content was considered to be clear and well presented; the amounts of information and language were deemed appropriate; and visual presentation of risk, such as for lymphoma (Figure 3), was considered particularly useful. Information presentation and the ability to tailor viewed content were considered to be well done; no major content modifications were recommended. The majority of the 15 clinicians (n=11) indicated that they would readily welcome myAID as a *positive addition* to their practice, 2 indicated that they would potentially use it, and the remaining 2 were uncertain. Those in favor of including myAID in their practice suggested that it could (1) improve time efficiency, as a result of patients continuing with the education in their own time; (2) deliver reliable information, which clinicians agree is accurate and evidence-based; (3) improve communication and therapeutic relationships; and (4) facilitate greater patient engagement and participation in SDM, potentially improving treatment adherence. Potential reasons for not routinely using myAID included (1) uncertainty of its benefits over carefully conducted face-to-face consultations, for example, for long-term patients with complex disease and treatment history; (2) feeling of *imposing* decisions on patients with potential for unnecessary anxiety or patient-clinician conflict; and (3) potential for increased resource and communication needs as a result of questions arising from using myAID.

Some clinicians suggested that there would be a potential for the DA to result in disagreement between the patient and clinician, which might impact the therapeutic relationship. In particular, it was expressed by some clinicians that the DA may provide patients with too much autonomy regarding their treatment decisions, the complexity of which might be better directed by their clinician, particularly in the setting of the software providing more generic and less tailored information to an individual patient's need.

Figure 3. Visual representation of risk of lymphoma (sample screenshot from myAID).

Perceived Role and Usefulness

Clinicians perceived that myAID would be useful as a long-term educational tool to supplement clinic visits and support decision-making for patients throughout their disease course. Clinicians suggested that one of the primary functions of DAs such as myAID would be to emphasize the importance of treatment adherence and the risk of the disease itself, and additionally, clinicians wanted to see further functions included that will enable monitoring and tracking of patient clinical information such as disease activity.

Target Population and Timing and Place of Use

Although all clinicians suggested that myAID, if used, should be used early in the disease course, it was highlighted that individual patients have differing information needs based on diagnosis recency, disease behavior, education, treatment history, and patient-clinician relationship. The majority of clinicians considered myAID to be of greatest benefit to patients at times of treatment escalation, specifically when considering treatments beyond 5-ASAs when trade-offs become more complex. Half the clinicians also considered use at diagnosis, although the remainder were concerned that this timing would provide patients with too much information, causing more anxiety than necessary, particularly if their disease was mild. There were differing views about its use in acutely unwell patients in the hospital setting among clinicians, some reporting potential utility as an educational tool providing a broad overview of all available treatment options for these patients, whereas others felt that it was less appropriate given the presence of other factors influencing treatment decisions and the time pressures in this context. Some clinicians felt that it was important that information about surgery was provided by clinicians initially. All clinicians felt that patients should access myAID in the privacy of their home, although some also perceived it to be beneficial to view it in the clinic before their consultation. It was universally agreed that myAID should be made accessible on mobile devices, particularly for patients with active disease and for patients to be reviewed either in the clinic or via phone consultation (gastroenterologist or IBD nurse) within 2 weeks of viewing myAID to facilitate decision-making and address questions promptly. However, all of the clinicians also highlighted the difficulty in scheduling such a visit and stated the importance of nursing follow-up support.

Potential Concerns and Areas for Improvement

In addition to issues already highlighted, clinicians identified the need to regularly update content (new information and drugs), citing the imminent arrival of biosimilars in Australia and potential future therapies including fecal microbiota transplantation. Further tailoring of information specific to the Australian environment was also suggested, such as greater emphasis on the risk of skin cancer for patients considering thiopurines. Additional support for patients, such as inclusion of links to patient support programs or organizations (eg, Crohn's and Colitis Australia), was also suggested. Although not identified as major issues, further improvements thought to be important included improved navigation between chapters (to allow for more rapid access to information of interest), enabling access in areas without internet connection, and ability to selectively print information included in myAID.

Discussion

Principal Findings

In this study, we have demonstrated that myAID, a web-based DA, is an acceptable and useful tool that can be used to facilitate SDM regarding medical and surgical treatments among patients with UC. Furthermore, using a participatory approach [13] in this pilot study, we obtained feedback from patients and clinicians that will enable us to optimize its usefulness and refine the clinical trial design used to evaluate its effectiveness.

The acceptability of DAs to facilitate SDM has been variable, and attrition rates associated with eHealth interventions have been high [12]. In our study, the majority of clinicians welcomed the prospect of SDM and use of the DA as an educational tool at times of treatment escalation, particularly when commencing immunosuppressive therapy to provide decision support and promote adherence. However, there were individual clinicians whose level of comfort using myAID varied depending on the clinical scenario to which it would be applied. Some clinicians did not perceive there to be any benefit from the DA over face-to-face consultations and suggested that the DA may be of limited value among patients with complex chronic disease, reinforcing the findings of Siegel et al [8] who found similar views expressed by gastroenterologists in relation to SDM. Moreover, some clinicians expressed greater risk of decisional conflict if patients felt obliged to make decisions using the DA, where previously their decisions would have been guided by the clinician, particularly at the time of diagnosis. These observations highlight the variation that is often observed in clinicians' approaches to patient education and practices. Although DAs have the potential to improve the quality of health care delivery, by minimizing variation in care, their capacity to do so is only as effective as their uptake by patients and their clinicians [18].

Patients often find the time restrictions and busyness of the outpatient clinic a difficult environment to obtain adequate information about their disease and treatment options, which often limits their ability to participate in SDM. In this study, the DA was accepted as a tool to supplement the provision of information by the majority of clinicians and valued by patients as an opportunity to engage in SDM. Although the majority of patients and clinicians found the content acceptable, there was a minority who expressed that the DA may provoke unnecessary anxiety, particularly among patients with a new diagnosis or mild disease. In particular, it was felt by some that DAs may be less appropriate for newly diagnosed patients who might benefit from a more tailored approach to patient education. Patients expressed satisfaction with the unbiased presentation of information and sense of control they had over the range and depth of information provided, which appeared to increase their sense of autonomy and enabled them to digest potentially confronting information at their own pace.

Previous studies have suggested that there is often discordance between what patients and clinicians prioritize and find useful in eHealth tools [8,11]. In this study, we identified differences between patients' and clinicians' perceptions regarding their role in management. In particular, patients expressed a desire to voice their preferences about treatment decisions, whereas clinicians perceived that one of the primary roles of myAID was to emphasize the importance of treatment adherence. This is in agreement with previous studies on eHealth technologies that have suggested that patients tend to prioritize convenience in contrast to clinicians who prioritize adherence [13]. Nonadherence remains a key barrier to the efficacy of medical treatment in UC, with rates up to 70% with 5-ASAs [19]; it has emerged that nonadherence does not relate to forgetting to take medication alone but may be voluntary, which may be attributed to patients' lack of belief in the value of their medical therapies [20,21]. This pilot study showed that myAID has the potential to converge patient and clinician perspectives by helping patients better understand the rationale for maintenance treatment and adherence and allowing for discussion of specific concerns or preferences. Existing literature has also previously suggested that DAs can reduce the discordance between patient and physician priorities [7].

Patients tend to overestimate the benefits of treatment and underestimate their harm [22]. SDM has been suggested as a strategy that may overcome this mismatch between patients' and clinicians' expectations [3]. In this study, information provided on surgical treatment and combination therapy was considered to be the most useful function of myAID by patients. In particular, patients expressed that the DA prompted them to openly discuss surgery with their clinicians and explore their values and preferences. Discussion about surgery is often neglected in routine UC management [23]. Although a DA has previously been developed for patients with UC to facilitate surgical decision-making between an end-ileostomy and ileal-pouch anal anastomosis [10], our DA was designed to place surgery in the context of medical treatment options at an earlier juncture in a patient's disease course. Early discussion of medical versus surgical treatment options facilitated by the DA serves the purpose of informing and reminding patients and their clinicians about the range of treatment options available, together with their associated risks versus benefits, which might help align patients' and clinicians' expectations.

Alterations in Study Design for the Planned CRCT

On the basis of the feedback obtained from this pilot study, specific alterations made to the study design for the planned CRCT included the exclusion of patients with newly diagnosed UC and those with acute severe UC requiring inpatient treatment. Drivers for these changes included concerns over heightened anxiety levels at the time of a new diagnosis and in the setting of an inpatient flare of disease, together with the time pressures associated with consultations and decision-making for both newly diagnosed and acutely unwell patients. The DA was considered to be more suitable for patients who were about to embark on immunosuppressive therapy, biologic therapy, or surgery, as an educational platform to help them consider the risk versus benefit of such management strategies. Given the preference expressed by patients to use the DA within the comfort of their own home, the study methods for the proposed CRCT were altered to provide all patients with a URL to allow them to use the DA from home. A scheduled study visit to follow-up within 2 weeks after the initial use of myAID was also introduced to help answer any questions or resolve any

decisional conflict. Further amendments to the content and structure of the DA will be considered based on the suggestions received in this pilot study and feedback resulting from the CRCT.

Limitations

To our knowledge, there are a number of DAs that have been developed to facilitate SDM for UC, each of which have different functions [12]. However, there is limited guidance as to what constitutes a good DA, as there are no universal measures available to determine optimal function except for one consensus checklist that documents quality criteria for development [14]. Although some DAs are designed to facilitate self-management via symptom monitoring and decision support [24,25], myAID's focus is on education as a strategy to increase patient participation in their treatment choices in an effort to increase patient engagement in their health care. Although we acknowledge that a limitation of our pilot study was that the sample size was small, the qualitative nature of our study is unique in that it represents one of the few studies that has adopted a participatory approach, involving both patients and clinicians, in the design and development of the DA as well as the design of the planned CRCT to evaluate its effectiveness. We further acknowledge that a limitation of myAID is that it is currently only available as a web-based medium via computer owing to its interactive component requiring the use of Flash media. However, although patients expressed a preference for mobile access to the DA, its desktop application was not identified as a major issue by the pilot study patients. The need for web literacy and English and Spanish language comprehension (in the Australian and US versions, respectively) are also limitations of the DA. However, these limitations are not felt to be insurmountable, as the platform used to create the DA can be customized for mobile device use and programmed for several other languages.

Conclusions and Future Directions—Unanswered Questions

The findings of this pilot study suggest that a web-based DA is an acceptable and useful tool to support decision-making regarding medical and surgical treatments among patients with UC. Using a participatory approach to engage patient and clinician feedback, the study design for the planned CRCT has been refined, thereby increasing the likelihood of being able to accurately evaluate whether myAID offers any benefit over usual care. Whether the use of DAs such as myAID that promote SDM will translate into clinically meaningful outcomes for patients remains to be seen and is the subject of our planned national CRCT.

Acknowledgments

The authors gratefully acknowledge the contributions of all patients, volunteers, psychologist (Afaf Girgis, Australia), nurse specialists (Alexandra Sechi, Australia and Elise Sawyer, Australia), and clinicians (Susan Connor, Australia; Watson Ng, Australia; and Jane Andrews, Australia) who helped with the Australianization of the UC decision aid. The original US version had the following key advisors: Corey Siegel, United States; Kimberly Thompson, United States; David Rubin, United States; Stefan Holubar, United States; Meenaskshi Bewtra, United States; Sonia Ramamoorthy, United States; Asher Kornbluth, United States; and Glyn Elwyn, United States. The authors gratefully acknowledge the support from Emmi Solutions who provided unrestricted access to the UC decision aid tool (myAID) to the authors and study participants during the study (Please note: they played no part in study design, data analysis, or manuscript preparation). Medibank Private (Australia) provided audio-recording for use in myAID in collaboration with Emmi Solutions. This study was supported by a South Western Sydney Local Health District Mid-Career Grant and funding provided by the Gastroenterological Society of Australia; PhD of AK is funded by the Australian Government Research Training Program Scholarship, and AG is supported by a Cancer Institute NSW grant. Centre for Oncology Education and Research Translation is funded by Cancer Institute NSW with contributions from the University of New South Wales; Western Sydney University; University of Wollongong, Illawarra, and Shoalhaven Local Health Districts; and the Ingham Institute for Applied Medical Research.

Conflicts of Interest

NK has received research support from Janssen. AJS is on advisory boards for Pfizer and has served as a consultant for AbbVie and Janssen. She has received educational support from Pfizer, Ferring, Orphan, Janssen, Abbvie Shire, and Takeda. WSWN had been on an advisory board for AbbVie and received research funding from AbbVie, Janssen, Takeda, Shire, and Aspen-Orphan. PDC has received educational support and has been a speaker at educational symposia sponsored by Janssen, AbbVie, Takeda, Ferring, Shire, and Baxter. PDC is on advisory boards for Janssen, Celgene, Takeda, Ferring, Shire, and Baxter. PDC was previously supported by a Gastroenterological Society of Australia Bushell Postdoctoral Fellowship and David Bickart Clinician Research Fellowship from the University of Melbourne and is currently supported by a National Medical Health & Research Council Early Career Fellowship. JMA is on advisory boards; has received speakers' fees and research support; and/or has coordinated education meetings for Abbott, AbbVie, Allergan, Bayer, Celegene, Ferring, Janssen, MSD, Pfizer, Shire, and Takeda. All the money provided to JMA have been received by her employer to support investigator-initiated research. CAS has served as a consultant for AbbVie, Amgen, Janssen, Lilly, Pfizer, Prometheus, Salix, and Takeda; has received speaker fees from AbbVie, Janssen, and Takeda; has received grant support from AbbVie, Janssen, Salix, and Takeda; and is the co-chair of the Crohn's and Colitis Foundation of America quality of care program. SJC is on advisory boards for AbbVie, Janssen, Pfizer, Ferring, Celgene, Takeda, and MSD. She has received speaker fees from AbbVie, Janssen, Shire, Ferring, Takeda, and Pfizer and educational

support from Shire and Takeda. The Liverpool Hospital IBD Service has received funding from AbbVie, Shire, Ferring, Takeda, and Orphan. The authors wish to state that, to the best of their knowledge, they do not believe these potential conflicts of interest have impacted any aspects of the research. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Interview guide for patients and clinicians.
[DOCX File , 17 KB - jmir_v23i2e15946_app1.docx]

References

1. Vaucher C, Maillard MH, Froehlich F, Burnand B, Michetti P, Pittet V. Patients and gastroenterologists' perceptions of treatments for inflammatory bowel diseases: do their perspectives match? Scand J Gastroenterol 2016 Sep;51(9):1056-1061. [doi: 10.3109/00365521.2016.1147065] [Medline: 26891800]
2. Peyrin-Biroulet L, Sandborn W, Sands BE, Reinisch W, Bemelman W, Bryant RV, et al. Selecting therapeutic targets in inflammatory bowel disease (STRIDE): determining therapeutic goals for treat-to-target. Am J Gastroenterol 2015 Sep;110(9):1324-1338. [doi: 10.1038/ajg.2015.233] [Medline: 26303131]
3. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
4. Siegel CA. Embracing the internet for progress in shared decision-making. Inflamm Bowel Dis 2007 Dec;13(12):1579-1580. [doi: 10.1002/ibd.20259] [Medline: 17879279]
5. Thompson KD, Connor SJ, Walls DM, Gollins J, Stewart SK, Bewtra M, et al. Patients with ulcerative colitis are more concerned about complications of their disease than side effects of medications. Inflamm Bowel Dis 2016 Apr;22(4):940-947. [doi: 10.1097/MIB.0000000000000740] [Medline: 26950308]
6. O'Connor A, Edwards, AG. The role of decision aids in promoting evidence-based patients choice. In: Evidence-Based Patient Choice. Oxford, United Kingdom: Oxford University Press; 2001:220-243.
7. Stacey D, Légaré F, Lewis KB. Patient decision aids to engage adults in treatment or screening decisions. J Am Med Assoc 2017 Aug 15;318(7):657-658. [doi: 10.1001/jama.2017.10289] [Medline: 28810006]
8. Siegel CA, Lofland JH, Naim A, Gollins J, Walls DM, Rudder LE, et al. Gastroenterologists' views of shared decision making for patients with inflammatory bowel disease. Dig Dis Sci 2015 Sep;60(9):2636-2645 [FREE Full text] [doi: 10.1007/s10620-015-3675-z] [Medline: 25939543]
9. Healthwise Staff. Ulcerative Colitis: Should I Have Surgery? HealthLink BC. 2019. URL: https://www.healthlinkbc.ca/health-topics/uf4785 [accessed 2017-12-01]
10. Cohan JN, Ozanne EM, Sewell JL, Hofer RK, Mahadevan U, Varma MG, et al. A novel decision aid for surgical patients with ulcerative colitis: results of a pilot study. Dis Colon Rectum 2016 Jun;59(6):520-528. [doi: 10.1097/DCR.0000000000000572] [Medline: 27145309]
11. Con D, Jackson B, Gray K, de Cruz P. eHealth for inflammatory bowel disease self-management - the patient perspective. Scand J Gastroenterol 2017 Sep;52(9):973-980. [doi: 10.1080/00365521.2017.1333625] [Medline: 28598210]
12. Jackson BD, Gray K, Knowles SR, de Cruz P. EHealth technologies in inflammatory bowel disease: a systematic review. J Crohns Colitis 2016 Sep;10(9):1103-1121. [doi: 10.1093/ecco-jcc/jjw059] [Medline: 26928960]
13. Jackson BD, Con D, De Cruz P. Design considerations for an eHealth decision support tool in inflammatory bowel disease self-management. Intern Med J 2018 Jun;48(6):674-681. [doi: 10.1111/imj.13677] [Medline: 29136332]
14. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, International Patient Decision Aids Standards (IPDAS) Collaboration. Developing a quality criteria framework for patient decision aids: online international delphi consensus process. Br Med J 2006 Aug 26;333(7565):417 [FREE Full text] [doi: 10.1136/bmj.38926.629329.AE] [Medline: 16908462]
15. Coulter A, Stilwell D, Kryworuchko J, Mullen PD, Ng CJ, van der Weijden T. A systematic development process for patient decision aids. BMC Med Inform Decis Mak 2013;13 Suppl 2:S2 [FREE Full text] [doi: 10.1186/1472-6947-13-S2-S2] [Medline: 24625093]
16. Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. Gut 1998 Jul;43(1):29-32 [FREE Full text] [doi: 10.1136/gut.43.1.29] [Medline: 9771402]
17. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
18. Jackson B, Begun J, Gray K, Churilov L, Liew D, Knowles S, et al. Clinical decision support improves quality of care in patients with ulcerative colitis. Aliment Pharmacol Ther 2019 Apr;49(8):1040-1051. [doi: 10.1111/apt.15209] [Medline: 30847962]
19. Vangeli E, Bakhshi S, Baker A, Fisher A, Bucknor D, Mrowietz U, et al. A systematic review of factors associated with non-adherence to treatment for immune-mediated inflammatory diseases. Adv Ther 2015 Nov;32(11):983-1028 [FREE Full text] [doi: 10.1007/s12325-015-0256-7] [Medline: 26547912]

20. Lenti MV, Selinger CP. Patients' beliefs and concerns about the necessity of long-term therapies are crucial for medication adherence for inflammatory bowel disease. J Crohns Colitis 2017 Jun 1;11(6):764-765. [doi: 10.1093/ecco-jcc/jjw213] [Medline: 27932452]
21. Selinger CP, Eaden J, Jones DB, Katelaris P, Chapman G, McDonald C, et al. Modifiable factors associated with nonadherence to maintenance medication for inflammatory bowel disease. Inflamm Bowel Dis 2013 Sep;19(10):2199-2206. [doi: 10.1097/MIB.0b013e31829ed8a6] [Medline: 23899547]
22. Hoffmann TC, Del Mar C. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. JAMA Intern Med 2015 Feb;175(2):274-286. [doi: 10.1001/jamainternmed.2014.6016] [Medline: 25531451]
23. Bewtra M, Kilambi V, Fairchild AO, Siegel CA, Lewis JD, Johnson FR. Patient preferences for surgical versus medical therapy for ulcerative colitis. Inflamm Bowel Dis 2014 Jan;20(1):103-114 [FREE Full text] [doi: 10.1097/01.MIB.0000437498.14804.50] [Medline: 24280881]
24. Elkjaer M, Shuhaibar M, Burisch J, Bailey Y, Scherfig H, Laugesen B, et al. E-health empowers patients with ulcerative colitis: a randomised controlled trial of the web-guided 'constant-care' approach. Gut 2010 Dec;59(12):1652-1661. [doi: 10.1136/gut.2010.220160] [Medline: 21071584]
25. Cross RK, Cheevers N, Rustgi A, Langenberg P, Finkelstein J. Randomized, controlled trial of home telemanagement in patients with ulcerative colitis (UC HAT). Inflamm Bowel Dis 2012 Jun;18(6):1018-1025 [FREE Full text] [doi: 10.1002/ibd.21795] [Medline: 21688350]

Abbreviations

5-ASA: 5-aminosalicylates
CRCT: cluster randomized controlled trial
DA: decision aid
IBD: inflammatory bowel disease
NSW: New South Wales
RA: research assistant
SCCAI: Simple Clinical Colitis Activity Index
SDM: shared decision making
UC: ulcerative colitis

Edited by G Eysenbach, Q Zeng; submitted 25.08.19; peer-reviewed by T Irizarry, C Hoving, G Baumblatt, D Con, K Carlsen; comments to author 02.09.20; revised version received 27.10.20; accepted 28.10.20; published 25.02.21.

Please cite as:
Kim AH, Girgis A, De Cruz P, Siegel CA, Karimi N, Ruban SO, Sechi AJ, Ng WSW, Andrews JM, Connor SJ
Development and Feasibility of a Web-Based Decision Aid for Patients With Ulcerative Colitis: Qualitative Pilot Study
J Med Internet Res 2021;23(2):e15946
URL: https://www.jmir.org/2021/2/e15946
doi:10.2196/15946
PMID:33629956

Original Paper

Husbands' Knowledge of Breast Cancer and Their Wives' Attitudes and Practices Related to Breast Cancer Screening in Saudi Arabia: Cross-sectional Online Survey

Afnan Abdulnasir Sabgul[1,2], MHA; Ameerah M N Qattan[2], PhD; Rubayyat Hashmi[3,4], MSc; Mohammed Khaled Al-Hanawi[2], PhD

[1]Academic Affairs and Training Department, King Abdulaziz Hospital, Makkah, Saudi Arabia

[2]Department of Health Services and Hospital Administration, Faculty of Economics and Administration, King Abdulaziz University, Jeddah, Saudi Arabia

[3]Centre for Health Research, University of Southern Queensland, Springfield, Australia

[4]School of Business, Faculty of Business, Education, Law and Arts, University of Southern Queensland, Toowoomba, Australia

Corresponding Author:
Mohammed Khaled Al-Hanawi, PhD
Department of Health Services and Hospital Administration
Faculty of Economics and Administration
King Abdulaziz University
University Building 125
Faculties Street
Jeddah, 80200
Saudi Arabia
Phone: 966 556522222
Email: mkalhanawi@kau.edu.sa

Abstract

Background: Despite Saudi Arabia's free and well-established cancer care program, breast cancer incidence and mortality are rising. Husbands' knowledge, and wives' attitudes and practices related to breast cancer screening are not well understood in Saudi Arabia.

Objective: The aim of this study was to investigate husbands' knowledge, and wives' attitudes and practices related to breast cancer screening in Saudi Arabia.

Methods: This cross-sectional study collected data from 403 husbands in the holy city of Makkah through an online self-reported questionnaire over a period of 2 months, from May 6 to July 7, 2020. Tabulation, bivariate, and multiple regression analyses were the major tools used for data analysis. Multivariate logistic regressions were used to examine the association between husbands' knowledge and wives' behavior regarding breast cancer screening methods.

Results: Husbands' knowledge score (a 1-point increase) was significantly associated with the wives' utilization of mammograms (adjusted odds ratio [AOR] 1.089, 95% CI 1.024-1.159) and breast self-examination (AOR 1.177, 95% CI 1.105-1.255). Husbands' knowledge also influenced the wives' attitudes toward learning about breast self-examination (AOR 1.138, 95% CI 1.084-1.195). There was no significant association between husbands' knowledge and wives' utilization of clinical breast examination. However, richer husbands showed a socioeconomic gradient concerning their wives' utilization of clinical breast examinations (AOR 2.603, 95% CI 1.269-5.341).

Conclusions: Overall, husbands' knowledge of breast cancer influences wives' attitudes and practices related to breast cancer screening methods in Saudi Arabia. Thus, interventions delivered to husbands might increase breast cancer awareness and survival.

(J Med Internet Res 2021;23(2):e25404) doi:10.2196/25404

KEYWORDS

attitude; breast cancer; husbands; knowledge; Saudi Arabia; screening

Introduction

Breast cancer is the leading cause of cancer-related deaths in women of over 100 countries, and is responsible for approximately 15% of all cancer deaths in women worldwide [1]. In the Kingdom of Saudi Arabia (KSA), breast cancer incidence increased by approximately 10-fold during the period of 1990-2016 [2]. Moreover, breast cancer at an advanced stage is more common in Saudi women compared to women from Western countries [3]. The incidence of breast cancer is also expected to show a rising trend in the next decade in the KSA due to the elderly age structure of the Saudi community [4,5]. If the current growth rate of breast cancer incidence in the KSA continues, it is expected to become the driving force in the projected 2015-2030 economic burden of US $159.44 billion for all cancers [6].

Breast cancer awareness contributes to the early detection of cancer presentation, and is thus necessary to reduce breast cancer incidence and mortality [7]. Evidence also suggests that a substantial proportion of breast cancer cases are self-detected [8]. Thus, high-risk groups need to be aware of breast cancer risk factors and risk reduction strategies to avoid being diagnosed in the late stages of cancer [9]. Knowledge about breast cancer can also promote utilization of the recommended screening methods for the timely detection of breast cancer [10,11]. However, women tend to have a limited level of knowledge about breast cancer in the KSA [12-14]. In addition, despite free access to health care services in the KSA [15], the level of breast cancer screening practices remains low [16]. Recent studies have also found marked socioeconomic gradients in breast cancer screening uptake among Saudi women [17].

Similar to several countries in the Middle East and the Arab world, the KSA has a conservative culture, and men usually manage many of the decisions, choices, and actions of women [18,19]. A sociocultural barrier exists for breast cancer screening practices, and failure to account for such barriers can undermine the success of well-established cancer care programs [20]. Although some studies have explored men's knowledge in relation to women's attitudes and practices related to breast cancer screening methods [18,19,21,22], these studies either did not explore the association between men's knowledge and their wives' practices regarding breast cancer screening or did not control for socioeconomic gradients. Thus, the degree to which men's knowledge can support women to take actions for better protection from breast cancer in a country such as the KSA is not fully understood. Understanding such a mechanism might help in developing breast cancer awareness interventions and reduce breast cancer mortality through early detection.

The aim of this study was to investigate the potential effect of husbands' knowledge on their wives' attitudes and practices related to breast cancer screening in the holy city of Makkah in the Makkah region of western Saudi Arabia, located 70 kilometers inland from Jeddah. Approximately 26% of the KSA population live in the Makkah region [23]. Makkah also hosts the largest annual Muslim pilgrimage, the Hajj, and is thus regarded as a holy city among Muslims. Moreover, the city has a diverse population, with different lifestyle patterns and economic statuses. To understand the impact of husbands' knowledge in Makkah city, we controlled for the possible effects of socioeconomic characteristics on wives' attitudes and practices related to breast cancer screening in this study.

Methods

Study Setting and Sample

This cross-sectional study was carried out in Makkah city of the Makkah region in the KSA from May 6 to July 7, 2020. Given the social/physical distancing measures implemented in the country because of the COVID-19 pandemic, data were collected via an online self-reported questionnaire using a Google Drive website. The link was distributed to potential participants via social media, including Twitter and WhatsApp groups.

This study included participants who were residents of Makkah city and were male partners (ie, husbands) aged between 20 and 80 years. According to the latest KSA census, Makkah has a total of 434,248 married men (ie, husbands) [23]. The minimum sample size necessary for the study was calculated to be 384 husbands (margin of error of 5%, confidence level of 95%, 50% response distribution, based on the total 434,248 husbands in the region).

The questionnaire was developed by the authors based on several previous studies [3,12,14,19,20]. The questionnaire consisted of three main parts: sociodemographic characteristics of the participants, knowledge about breast cancer and screening methods, and knowledge of wives' practices and attitudes related to breast cancer screening. The questionnaire was written in Arabic. It was initially drafted by two authors (AS and MA) in English and then translated to Arabic by AQ. The questionnaire was also back-translated to English by two bilingual speakers, and reviewed by academic researchers from the Department of Health Services and Hospital Administration at King Abdulaziz University to maintain the original construct and to ensure the meaning of the content.

On the first page of the online questionnaire, participants were provided with the study background, aim, and objectives. They were informed that they did not have to complete the questionnaire and were free to withdraw at any point without providing a reason. They were also informed that there would be no sensitive personal questions, and all information provided would be anonymous, confidential, and used for research purposes only. Participants aged 20 years or over living in Makkah city, who understood the content of the questionnaire, and agreed to participate in the study were instructed to start the questionnaire. Online informed consent was obtained from all participants before proceeding with the questionnaire.

Variables

In this study, we defined four outcome variables: variables related to wives' practices of breast self-examination (BSE) (yes, no, and unsure), clinical breast examination (CBE) (yes, no, and unsure), and mammograms (yes, no, and unsure), and a variable related to wives' attitudes toward learning about BSE (yes, no, and unsure). The main exposure variable of this study was the husbands' knowledge scores. Using predetermined

answer keys based on medical evidence, each correctly answered knowledge question was awarded one knowledge point. Incorrectly answered questions or "I don't know" responses were awarded zero knowledge points.

There were 20 knowledge questions in the questionnaire, including two questions on BSE (age of commencement and frequency), one question on CBE (age of commencement), two questions on mammograms (age of commencement and frequency), five questions on symptoms (changes in breast size, nipple morphology, nipple dryness, nipple discharge, and lymphadenopathy), six questions on risk factors (use of contraceptives, hormone replacement therapy, overexposure to radiation, smoking, hereditary, and old age), and four questions on protective factors (natural breastfeeding, sports/exercise practices, early conception, and balanced and healthy diet). The total knowledge score was computed by adding up the points for each of the 20 breast cancer knowledge questions, yielding a total breast cancer knowledge score between 0 and 20.

Multivariate logistic regression analyses were performed controlling for the husbands' sociodemographic covariates. We used age, nationality, education, occupation, and household monthly income level as the sociodemographic control variables. Age was categorized as 20-29 (reference category), 30-39, 40-49, 50-59, and ≥60 years. Nationality was categorized as non-Saudi (reference category) and Saudi. Education was categorized as high school education level or less (reference category) and college/university degree or higher. Occupation was split into the categories of unemployed or retired (reference category), self-employed, government employee, and private sector employee. Household monthly income was grouped into two categories: less than 10,000 Saudi Riyal (SR) (reference category) and ≥10,000 SR (1 SR=US $0.27).

Statistical Analyses

Tabulation, bivariate, and multiple regression analyses were the major data analytic tools used in this study. Means (SD) are used to describe the continuous variables, whereas frequency and percentages are used to describe the categorical variables. The nonparametric univariate χ^2 goodness-of-fit test was used to assess the statistical significance of the husbands' indicators of knowledge on breast cancer signs and preventive measures. Multivariate logistic binary regression analysis was used to assess the correlations between the husbands' knowledge about breast cancer and their sociodemographic, educational, and economic factors with their wives' various types of breast screening behaviors. The associations are expressed as adjusted odds ratios with 95% CIs. The SPSS version 20 software package was used to perform all statistical analyses.

Ethical Clearance

All procedures performed in this study involving human participants were performed in accordance with the ethical standards of the institutional or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This study was reviewed and approved by the King Abdulaziz University Research Ethics Committee, and was designed and performed in accordance with the ethical principles established by the university. This study also received ethical approval from the Ministry of Health of the KSA (IRB no: H-02-K-076-0320-275). Online informed consent to participate was secured from all respondents who participated in this study. The data collection procedure was anonymous and as such no personal identifying information was collected. The data were stored on the first author's password-protected personal computer. An additional copy of the data was stored on a password-protected external drive kept by the first author to serve as a backup.

Results

Demographic and Socioeconomic Characteristics of the Respondents

A total of 445 participants completed the questionnaire. After excluding 42 respondents who reported living outside of Makkah city, the final sample consisted of 403 participants. Table 1 shows the diverse demographic and socioeconomic characteristics of the study participants.

The majority of participants were Saudi citizens, and the others comprised expatriate men residing in Makkah city. Of the 403 participants, over half were aged 40 or older, over three-quarters had a college/university or a postgraduate degree, and just over half worked in various governmental jobs. Over one-third of participants had a household monthly income of less than 10,000 SR.

Table 1. Demographic and socioeconomic characteristics (N=403).

Characteristic	Frequency, n (%)	95% CI
Age (years)		
20-29	64 (15.9)	12.7-19.1
30-39	120 (29.8)	25.3-34.2
40-49	106 (26.3)	22.1-30.3
50-59	62 (15.4)	12.4-18.6
≥60	51 (12.7)	10.2-14.7
Nationality		
Non-Saudi	56 (13.9)	10.7-17.1
Saudi	347 (86.1)	82.9-89.3
Educational level		
High school level or below	85 (21.1)	17.4-24.6
College/university degree	248 (61.5)	57.3-66.2
Postgraduate degree	70 (17.4)	14.1-20.6
Occupation		
Retired/unemployed	79 (19.6)	16.4-22.6
Self-employed	19 (4.7)	3.0-6.7
Governmental employee	214 (53.1)	48.6-57.8
Private sector employee	91 (22.6)	19.1-26.8
Household monthly income (Saudi Riyal)[a]		
5000	39 (9.7)	7.2-12.4
5000 to <7000	37 (9.2)	6.7-11.9
7000 to <10,000	72 (17.9)	14.6-21.1
10,000 to <15,000	142 (35.2)	31.3-39.5
15,000 to <20,000	57 (14.1)	11.2-17.4
20,000 to <30,000	42 (10.4)	7.9-13.4
≥30,000	14 (3.5)	2.2-5.2

[a]1 Saudi Riyal=US $0.27.

Husbands' Knowledge and Awareness of Female Breast Cancer

Table 2 shows the husbands' knowledge and awareness of breast cancer screening methods, including BSE, CBE, and mammograms.

In general, the basic knowledge and awareness of female BSE, CBE, and mammograms was low among the husbands in Makkah city. The majority of the husbands had heard about female BSE; however, when asked about the stage at which women should start BSE, just over a quarter of the respondents correctly inferred that they should start from the age of 20. The same proportion also correctly suggested that monthly BSE is recommended for women. In contrast, over half of the respondents had not heard about CBE, and only about 35% of the husbands correctly answered that the recommended time of CBE is from the age of 40 or when women have symptoms. Furthermore, approximately 60% of the husbands had heard about mammograms. However, less than half of the respondents correctly inferred that the recommended starting age of mammograms is from the age of 40 or higher. Similarly, only about a third of the husbands correctly suggested that the recommended frequency of mammograms for women is every 1-2 years.

The husbands' poor knowledge of breast cancer screening was also reflected in their wives' attitudes and practices related to these behaviors. As shown in Table 3, according to the husbands, only approximately 20% of the wives regularly self-examined their breasts or had ever had a CBE. Similarly, less than 20% of the wives had ever had a mammogram. However, over half of the husbands stated that their wives were willing to learn more about BSE.

The results related to the husbands' knowledge of female breast cancer symptoms, risk factors, and preventive measures are shown in Table 4.

Table 2. Husbands' knowledge and awareness of breast self-exams (BSE), clinical breast exams (CBE), and mammograms (N=403).

Question	Frequency, n (%)	95% CI
BSE		
Have you ever heard of female BSE?		
No	107 (26.6)	22.8-31.0
Yes	296 (73.4)	69.0-77.4
At what stage do you think women should start BSE?		
Do not know	145 (36.0)	31.5-40.4
At puberty	50 (12.4)	9.4-15.5
From the age of 20	108 (26.8)	22.6-31.0
From the age of 40	92 (22.8)	19.1-26.8
After menopause	8 (2.0)	1.0-3.0
How often should women undergo BSE?		
Do not know	170 (42.2)	37.7-46.7
Weekly	26 (6.5)	4.5-8.4
Monthly	106 (26.3)	22.3-30.2
Annually	101 (25.1)	21.1-29.3
CBE		
Have you ever heard of female CBE?		
No	229 (56.8)	52.1-61.8
Yes	174 (43.2)	38.7-47.6
At what age do you think women should start CBE?		
Do not know	220 (54.6)	49.6-59.3
At puberty	12 (3.0)	1.5-4.6
From the age of 20 or higher	29 (7.2)	5.2-9.2
From the age of 40 or when there are symptoms	139 (34.5)	30.0-39.0
After menopause	3 (0.7)	0.0-1.7
Mammogram		
Have you ever heard of a mammogram?		
No	161 (40.0)	35.2-44.4
Yes	242 (60.0)	55.6-64.5
At what age do you think women should start having mammograms?		
Do not know	199 (49.4)	44.9-54.1
At puberty	3 (0.4)	0.0-1.7
From the age of 20 or higher	20 (5.0)	3.2-6.9
From the age of 40 or higher	169 (41.9)	37.5-46.2
After menopause	12 (3.0)	1.7-4.2
How often should women undergo a mammogram?		
Do not know	210 (52.1)	47.4-56.5
Monthly	12 (3.0)	1.7-4.5
Every 1-2 years	145 (36.0)	31.5-40.7
Every 4 years	36 (8.9)	6.5-11.7

Table 3. Husbands' knowledge of their wives' practices and attitudes related to breast cancer screening (N=403).

Outcome variables	Frequency, n (%)	95% CI
Has your wife ever had a mammogram?		
No	334 (82.9)	79.4-86.4
Yes	69 (17.1)	14.1-20.6
Has your wife ever had a clinical breast exam?		
No	322 (79.9)	75.9-83.4
Yes	81 (20.1)	16.6-23.6
Does your wife self-examine her own breasts regularly?		
No	328 (81.4)	77.9-84.9
Yes	75 (18.6)	15.1-22.1
Do you think your wife is willing to learn more about breast self-examination?		
No	180 (44.7)	40-49.6
Yes	223 (55.3)	51.1-59.8

Table 4. Husbands' knowledge about female breast cancer symptoms, risk factors, and preventive measures (N=403).

Question	Incorrect answer		Correct answer		GOF[a] χ^2 (*df*=2)	*P* value
	n (%)	95% CI	n (%)	95% CI		
Symptoms						
The change in women's breast size is a sign of breast cancer	285 (70.7)	66.7-75.2	118 (29.3)	25.1-33.0	106.784	<.001
Lymphadenopathy is a sign/symptom of breast cancer	162 (40.4)	35.7-44.7	241 (59.6)	55.3-64.3	170.462	<.001
Nipple skin dryness and peeling is a sign of breast cancer	299 (74.2)	70.2-78.2	104 (25.8)	22.0-29.8	175.985	<.001
Nipple discharge is a sign of breast cancer	273 (67.7)	63.5-72.0	130 (32.3)	27.8-36.7	172.263	<.001
Changes in nipple morphology is a sign of breast cancer	268 (66.5)	61.8-71.1	135 (33.5)	29.0-38.0	167.29	<.001
Risk factors						
Use of contraceptives is a possible risk factor for breast cancer	333 (82.6)	78.9-85.9	70 (17.4)	14.4-20.8	150.02	<.001
Hormonal replacement therapy may predispose women to breast cancer	338 (83.9)	80.1-87.1	65 (16.1)	13.2-19.6	275.275	<.001
Overexposure to radiation may predispose women to breast cancer	160 (39.7)	35.2-44.7	243 (60.3)	55.6-64.5	209.037	<.001
Cigarette smoking is a risk factor for breast cancer among women	144 (35.7)	31.0-40.2	259 (64.3)	60.3-68.7	195.042	<.001
Women may develop breast cancer due to hereditary factors	123 (30.5)	26.1-35.0	280 (69.5)	65.3-73.9	260.164	<.001
Older age is a factor that enhances the chance of developing breast cancer among women	259 (64.3)	59.6-68.7	144 (35.7)	31.5-40.0	59.201	<.001
Protective factors						
Natural breastfeeding may prevent against developing breast cancer among women	103 (25.6)	21.8-29.8	300 (74.4)	70.5-78.2	326.298	<.001
Regular sports/exercise may protect against breast cancer	128 (31.8)	27.4-36.3	275 (68.2)	63.8-72.2	258.169	<.001
Early conception before the age of 30 years may reduce the chance of getting cancer among women	325 (80.6)	77.2-84.6	78 (19.4)	15.4-22.8	258.854	<.001
A balanced and healthy diet may reduce the chance of developing breast cancer among women	123 (30.5)	26.3-35.7	280 (69.5)	65.3-73.7	277.97	<.001

[a]GOF: goodness of fit.

The most frequent symptom of breast cancer that had been correctly identified was the sign of lymphadenopathy, followed by changes in nipple morphology, nipple discharge, changes in breast size, and nipple skin dryness. Similarly, the most frequent

risk factor of breast cancer mentioned by husbands was hereditary factors, followed by cigarette smoking, overexposure to radiation, older age, use of contraceptives, and hormonal replacement therapy. In addition, the most frequent protective factor of breast cancer known to husbands was natural breastfeeding, followed by a balanced and healthy diet, regular exercise, and early conception before the age of 30.

The knowledge score of the husbands was calculated from the scores of the 20 knowledge questions shown in Tables 2 and 4. The score ranged from 0 to 20, with a mean knowledge score of 8.31 points (SD 4.91). Those with a lower knowledge score were regarded as having poorer knowledge of breast cancer compared to those with a higher knowledge score. The knowledge score was then used to investigate any effects of the husbands' knowledge on their wives' attitudes and practices related to breast cancer screening. The details of the regression analyses are discussed in the following section.

Regression Analyses

Table 5 shows the results of the multivariate logistics regression analyses used to assess the adjusted association of husbands' sociodemographic characteristics and knowledge about breast cancer with their wives' attitudes and practices related to breast cancer screening. Husband age of 40-49, 50-59, and ≥60 years was significantly associated with their wives' practices of mammogram screening. However, only age of 40-49 years was significantly associated with their wives' practices of CBE. Similarly, a husband age of 30-39 and 40-49 years was associated with their wives' practices of BSE. Only older husbands aged 60 years or more had significantly lower odds in terms of their wives' willingness to learn more about BSE.

The results also revealed that husbands who were employed in the private sector had more than a 4-times higher odds that their wives had performed mammography screening than those who were retired or unemployed. Similarly, the husbands with higher household monthly income levels (≥10,000 SR) had approximately 2-times higher odds that their wife had undergone a mammogram and had more than 2-times higher odds that their wives had undergone a CBE. Additionally, the husbands' high knowledge score was significantly associated with their wives' utilization of mammograms and BSE, as well as their wives' willingness to learn more about BSE. However, no association between husbands' knowledge scores and wives' utilization of CBE screening was found (Table 5).

Table 5. Multivariate logistic regression analyses of the effects of husbands' knowledge on their wives' attitudes and practices related to breast cancer screening.

Variables	Wife underwent a mammogram, AOR[a] (95% CI)	Wife underwent a CBE[b], AOR (95% CI)	Wife underwent a BSE[c], AOR (95% CI)	Wife willing to learn more about BSE, AOR (95% CI)
Age (years)				
30-39	2.084 (0.638-6.812)	2.080 (0.781-5.54)	4.595 (1.482-14.246)***	1.457 (0.761-2.788)
40-49	4.332 (1.372-13.675)**	3.207 (1.22-8.425)**	4.363 (1.352-14.079)**	1.691 (0.862-3.318)
50-59	5.34 (1.423-20.10)**	2.764 (0.897-8.515)	2.852 (0.722-11.263)	1.188 (0.522-2.705)
≥60	6.54 (1.325-32.56)**	1.880 (0.419-8.432)	2.414 (0.473-12.329)	0.220 (0.069-0.698)**
Saudi nationality	1.698 (0.639-4.513)	1.305 (0.495-3.437)	2.030 (0.726-5.677)	2.091 (0.954-4.584)
College/university degree or higher	0.772 (0.355-1.67)	0.665 (0.323-1.381)	1.557 (0.641-3.780)	0.549 (0.296-1.018)
Occupation				
Self-employed	2.109 (0.398-11.183)	0.327 (0.039-3.588)	1.101 (0.177-5.639)	0.599 (0.167-2.15)
Government employee	0.646 (0.211-1.980)	0.997 (0.344-3.588)	0.387 (0.122-1.231)	0.644 (0.271-1.527)
Private sector employee	4.735 (1.319-16.990)**	2.677 (0.789-9.084)	1.394 (0.401-4.844)	1.097 (0.409-2.939)
Household monthly income ≥10,000 SR[d]	1.965 (0.925-4.174)	2.603 (1.269-5.341)***	1.222 (0.602-2.479)	0.710 (0.411-1.228)
Knowledge score	1.089 (1.024-1.159)***	1.045 (0.989-1.108)	1.177 (1.105-1.255)***	1.138 (1.084-1.195)***

[a]AOR: adjusted odds ratio.

[b]CBE: clinical breast examination.

[c]BSE: breast self- examination.

[d]SR: Saudi Riyal (1 SR=US $0.27).

*P<.10, **P<.05, and ***P<.01.

Discussion

Principal Findings

Although previous studies have addressed the knowledge and attitude of breast cancer screening among Saudi women, this study illustrates the impact of husbands' knowledge on their wives' breast cancer screening behavior in Makkah city of the KSA using an online cross-sectional survey. The findings revealed that husbands' knowledge has an important impact on their wives' attitudes and practices related to the utilization of breast cancer screening methods. This study also showed

socioeconomic gradients concerning wives' utilization of mammogram and CBE.

Breast cancer survival is closely associated with cancer awareness [7,24]. Because of the conservative culture in the KSA, many women refrain from seeking medical advice until breast cancer presents at an advanced stage [25]. Moreover, some women cannot have any screening of their breasts without first obtaining approval from their male guardians [20]. Thus, men have an important role in supporting and encouraging women to obtain early diagnosis and treatment. However, the magnitude of the influence that husbands' knowledge of breast cancer has on their wives' screening behavior was unclear in earlier studies. Our findings have strengths that might be relevant in addressing the information needs of breast cancer patients [26] or implementing risk-stratified breast cancer prevention strategies [27].

One of the major findings of this study was that the husbands' knowledge was significantly associated with their wives' utilization of mammograms and BSE. This study also found that a 1-point increase in the husbands' knowledge scores on breast cancer increased the odds of mammogram utilization by approximately 1.1 times and increased the odds of BSE utilization by approximately 1.2 times. In addition, a 1-point increase in the husbands' knowledge scores on breast cancer increased the odds of their wives' being willing to learn about BSE by approximately 1.1 times. If considered cumulatively, our findings suggest that the husbands' knowledge substantially increased the probability of their wives' participation in mammogram and BSE screening methods. Similar findings were also reported in another study [19].

Most previous studies have not controlled for all socioeconomic factors when investigating husbands' knowledge about breast cancer [18-22]. However, several other studies found socioeconomic gradients in breast cancer incidence, mortality, and screening uptake behavior [17,28,29]. In our study, the findings of socioeconomic gradients were mixed, showing limited socioeconomic gradients in terms of mammogram (private sector employee) and CBE (income level) uptake. However, this study also found no association between education, nationality, and screening practices. This could be due to our small sample size. Despite the majority of our sample participants being highly educated, the level of breast cancer knowledge among the husbands in Makkah city was not satisfactory. Similar findings of men's poor knowledge about breast cancer were also reported in the Jeddah [21,22], Hail [3], and Asir [14,19] regions of Saudi Arabia, as well as in other countries such as Jordan [18] and Ghana [30].

In line with existing evidence [19], our study also found a significant association between the age of the husbands and their wives' utilization of mammograms, CBEs, and BSE, as well as their wives' attitudes toward learning about BSE. We found that only a husband age of 40-49 years had a significant relationship with CBE uptake, whereas almost all ages (40-49, 50-59, and 60 years or older) showed a significant association with mammogram uptake. BSE was usually performed by the wives of younger husbands (30-39 and 40-49 years), whereas the wives of elderly husbands (60 years or older) were less willing to learn about BSE. Our findings suggest that the age distribution of the husbands had a significant impact on their wives' utilization of screening methods and attitudes toward breast cancer.

In the KSA, a large number of breast cancer cases are discovered at late stages, which leads to a lower rate of recovery [31]. Despite free health services and free screening facilities, there is almost no use of mammograms in the KSA [15,16]. Thus, increasing the awareness about early detection tools for breast cancer is vitally important. Recent studies have found that breast cancer awareness interventions (eg, public health campaigns and educational programs) are effective to increase the uptake of BSE and mammogram behaviors, and to increase the probability of breast cancer screening attendance [32]. Studies have also found that target-based interventions can increase breast cancer awareness [9]. Our study findings indicate that there is a lot of work that needs to be done to improve cancer survival rates and to lower incidence rates. It is strongly recommended to empower women, raise cancer awareness, and remove barriers to fight breast cancer.

Limitations

The findings of this study should be interpreted in light of some limitations. This research concerned the Makkah region and urban settings only, where the majority of the men were well educated. In addition, the study design was cross-sectional in nature with the possibility of recall and self-reporting biases. Finally, the descriptive analysis cannot be generalized to all husbands in the KSA, as study participants were recruited through a convenience sampling method. However, care was taken to recruit participants that preserve the sociodemographic distribution of the KSA and, as already discussed, our descriptive analysis did not contradict the findings of past studies.

Conclusions

This study found evidence that husbands' knowledge about breast cancer has an important role in encouraging the breast cancer screening behaviors of their wives. Furthermore, the study findings documented limited knowledge about breast cancer among Makkah residents. Moreover, the wives of elderly husbands, according to their husbands' beliefs, are less willing to learn about BSE, whereas younger husbands' wives had higher odds of performing BSE. We also found some limited socioeconomic gradients regarding the utilization of mammograms and CBEs. These findings have important implications for conservative cultures where men play a crucial role in women's decisions and actions. Policies could be directed to raise breast cancer awareness, and interventions could be devised such that high-risk groups (eg, elderly or low socioeconomic status background) adhere to the recommended cancer screening protocol. Efforts should be made to empower women so that barriers to seeking medical help are alleviated. Future research is needed to evaluate whether implementing breast cancer awareness interventions increases cancer survival.

Authors' Contributions

Conceptualization: AS and MA; methodology: AS and MA; software: AS; validation: AS, MA, and AQ; formal analysis: AS; investigation: AS; data curation: AS and MA; writing—original draft preparation: AS, AQ, RH, and MA; writing—review and editing: AS, AQ, RH, and MA; supervision: MA and AQ; project administration: AS, MA, and AQ. All authors read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424. [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Althubiti MA, Nour Eldein MM. Trends in the incidence and mortality of cancer in Saudi Arabia. Saudi Med J 2018 Dec;39(12):1259-1262. [doi: 10.15537/smj.2018.12.23348] [Medline: 30520511]
3. Alharbi SH, Alreshidi FS, Ahmed IAB, Alrashidi AG, Alrashedi SA, Alshammeri KJK, et al. Assessment of knowledge and perception towards breast cancer prevention and early detection. Int J Med Res Health Sci 2018;7(1):65-76 [FREE Full text]
4. Al-Suroj HR, Al Ali FB, Al-Saleh KH. Awareness and attitude among Saudi females toward breast cancer screening in Al-Ahsa, KSA. Egypt J Hosp Med 2018 Apr;71(2):2516-2522. [doi: 10.12816/0045650]
5. Ibrahim EM, Zeeneldin AA, Sadiq BB, Ezzat AA. The present and the future of breast cancer burden in the Kingdom of Saudi Arabia. Med Oncol 2008;25(4):387-393. [doi: 10.1007/s12032-008-9051-5] [Medline: 18317955]
6. Jazieh AR, Da'ar OB, Alkaiyat M, Zaatreh YA, Saad AA, Bustami R, et al. Cancer incidence trends from 1999 to 2015 and contributions of various cancer types to the overall burden: Projections to 2030 and extrapolation of economic burden in Saudi Arabia. Cancer Manag Res 2019;11:9665-9674. [doi: 10.2147/CMAR.S222667] [Medline: 32009819]
7. Austoker J, Bankhead C, Forbes LJL, Atkins L, Martin F, Robb K, et al. Interventions to promote cancer awareness and early presentation: systematic review. Br J Cancer 2009 Dec 03;101(Suppl 2):S31-S39 [FREE Full text] [doi: 10.1038/sj.bjc.6605388] [Medline: 19956160]
8. Roth MY, Elmore JG, Yi-Frazier JP, Reisch LM, Oster NV, Miglioretti DL. Self-detection remains a key method of breast cancer detection for U.S. women. J Womens Health (Larchmt) 2011 Aug;20(8):1135-1139 [FREE Full text] [doi: 10.1089/jwh.2010.2493] [Medline: 21675875]
9. Linsell L, Forbes LJL, Kapari M, Burgess C, Omar L, Tucker L, et al. A randomised controlled trial of an intervention to promote early presentation of breast cancer in older women: effect on breast cancer awareness. Br J Cancer 2009 Dec 03;101(Suppl 2):S40-S48 [FREE Full text] [doi: 10.1038/sj.bjc.6605389] [Medline: 19956161]
10. Allen TL, Van Groningen BJ, Barksdale DJ, McCarthy R. The breast self-examination controversy: what providers and patients should know. J Nurse Practit 2010 Jun;6(6):444-451. [doi: 10.1016/j.nurpra.2009.11.005]
11. Oeffinger KC, Fontham ETH, Etzioni R, Herzig A, Michaelson JS, Shih YT, American Cancer Society. Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. JAMA 2015 Oct 20;314(15):1599-1614 [FREE Full text] [doi: 10.1001/jama.2015.12783] [Medline: 26501536]
12. Sait WA, Al-Amoudi SM, Tawatai DA, Abduljabbar HS. The knowledge of breast cancer among young Saudi females. Saudi Med J 2010;31(11):179-181.
13. Abdel-Aziz SB, Amin TT, Al-Gadeeb MB, Alhassar AI, Al-Ramadan A, Al-Helal M, et al. Perceived barriers to breast cancer screening among Saudi women at primary care setting. Asian Pac J Cancer Prev 2017 Sep 27;18(9):2409-2417 [FREE Full text] [doi: 10.22034/APJCP.2017.18.9.2409] [Medline: 28950697]
14. Alshahrani SM, Fayi KA, Alshahrani SH, Alahmari DS, Al Bejadi KM, Alahmari DM, et al. Assessment awareness of public about breast cancer and its screening measurements in Asir Region, KSA. Indian J Surg Oncol 2019 Jun;10(2):357-363 [FREE Full text] [doi: 10.1007/s13193-019-00899-5] [Medline: 31168262]
15. Gosadi IM. National screening programs in Saudi Arabia: Overview, outcomes, and effectiveness. J Infect Public Health 2019;12(5):608-614 [FREE Full text] [doi: 10.1016/j.jiph.2019.06.001] [Medline: 31248815]
16. El Bcheraoui C, Basulaiman M, Wilson S, Daoud F, Tuffaha M, AlMazroa MA, et al. Breast cancer screening in Saudi Arabia: free but almost no takers. PLoS One 2015;10(3):e0119051 [FREE Full text] [doi: 10.1371/journal.pone.0119051] [Medline: 25774520]
17. Al-Hanawi MK, Hashmi R, Almubark S, Qattan AMN, Pulok MH. Socioeconomic inequalities in uptake of breast cancer screening among Saudi women: a cross-sectional analysis of a national survey. Int J Environ Res Public Health 2020 Mar 20;17(6):2056 [FREE Full text] [doi: 10.3390/ijerph17062056] [Medline: 32244881]
18. Al Dasoqi K, Zeilani R, Bawadi H, Al Dasoqi A. Perspectives and attitudes of Jordanian male college students on breast cancer screening. J Cancer Educ 2017 Mar;32(1):24-30. [doi: 10.1007/s13187-015-0909-x] [Medline: 26386596]

19. Al-Musa HM, Awadalla NJ, Mahfouz AA. Male partners' knowledge, attitudes, and perception of women's breast cancer in Abha, Southwestern Saudi Arabia. Int J Environ Res Public Health 2019 Aug 25;16(17):3089 [FREE Full text] [doi: 10.3390/ijerph16173089] [Medline: 31450695]
20. Al-Amoudi SM, Abduljabbar HS. Men's knowledge and attitude towards breast cancer in Saudi Arabia. A cross-sectional study. Saudi Med J 2012 May;33(5):547-550. [Medline: 22588817]
21. Al-Amoudi S, Eldeek B, Ayuob N, Alzhrani W, Alahwal M. Knowledge about breast cancer among male medical students, Jeddah, 2011. Life Sci J 2012;9(3):166-170.
22. Al-Amoudi S, AlHomied MTA, AlSayegh NYN, Radi ONI, Zagzoog MMS, Aloufi OFM, et al. Breast cancer knowledge among male high school students in Saudi Arabia. J Cancer Educ 2016 Dec;31(4):784-788. [doi: 10.1007/s13187-015-0951-8] [Medline: 26611278]
23. Statistical Yearbook of 2018. General Authority for Statistics, Kingdom of Saudi Arabia. 2018. URL: https://www.stats.gov.sa/en/258 [accessed 2020-09-02]
24. Niksic M, Rachet B, Duffy SW, Quaresma M, Møller H, Forbes LJ. Is cancer survival associated with cancer symptom awareness and barriers to seeking medical help in England? An ecological study. Br J Cancer 2016 Sep 27;115(7):876-886 [FREE Full text] [doi: 10.1038/bjc.2016.246] [Medline: 27537388]
25. Mahfouz AA, Hassanein MHA, Nahar S, Farheen A, Gaballah II, Mohamed A, et al. Breast cancer knowledge and related behaviors among women in Abha City, southwestern Saudi Arabia. J Cancer Educ 2013 Sep;28(3):516-520. [doi: 10.1007/s13187-013-0495-8] [Medline: 23797712]
26. Lu H, Xie J, Gerido LH, Cheng Y, Chen Y, Sun L. Information needs of breast cancer patients: theory-generating meta-synthesis. J Med Internet Res 2020 Jul 28;22(7):e17907 [FREE Full text] [doi: 10.2196/17907] [Medline: 32720899]
27. Kukafka R, Yi H, Xiao T, Thomas P, Aguirre A, Smalletz C, et al. Why breast cancer risk by the numbers is not enough: evaluation of a decision aid in multi-ethnic, low-numerate women. J Med Internet Res 2015 Jul 14;17(7):e165 [FREE Full text] [doi: 10.2196/jmir.4028] [Medline: 26175193]
28. Palència L, Espelt A, Rodríguez-Sanz M, Puigpinós R, Pons-Vigués M, Pasarín MI, et al. Socio-economic inequalities in breast and cervical cancer screening practices in Europe: influence of the type of screening program. Int J Epidemiol 2010 Jun;39(3):757-765. [doi: 10.1093/ije/dyq003] [Medline: 20176587]
29. Kelly DM, Estaquio C, Léon C, Arwidson P, Nabi H. Temporal trend in socioeconomic inequalities in the uptake of cancer screening programmes in France between 2005 and 2010: results from the Cancer Barometer surveys. BMJ Open 2017 Dec 14;7(12):e016941 [FREE Full text] [doi: 10.1136/bmjopen-2017-016941] [Medline: 29247085]
30. Ameade EPK, Amalba A, Kudjo T, Kumah MK, Mohammed BS. Reducing the breast cancer menace: the role of the male partner in Ghana. Asian Pac J Cancer Prev 2014;15(19):8115-8119 [FREE Full text] [doi: 10.7314/apjcp.2014.15.19.8115] [Medline: 25338993]
31. Alotaibi RM, Rezk HR, Juliana CI, Guure C. Breast cancer mortality in Saudi Arabia: Modelling observed and unobserved factors. PLoS One 2018;13(10):e0206148 [FREE Full text] [doi: 10.1371/journal.pone.0206148] [Medline: 30347002]
32. Anastasi N, Lusher J. The impact of breast cancer awareness interventions on breast screening uptake among women in the United Kingdom: A systematic review. J Health Psychol 2019 Jan;24(1):113-124. [doi: 10.1177/1359105317697812] [Medline: 28810435]

Abbreviations

BSE: breast self-examination
CBE: clinical breast examination
KSA: Kingdom of Saudi Arabia
SR: Saudi Riyal

Edited by R Kukafka; submitted 31.10.20; peer-reviewed by C Jacob, AJ Nagarajan; comments to author 24.12.20; revised version received 29.12.20; accepted 31.01.21; published 25.02.21.

Please cite as:
Sabgul AA, Qattan AMN, Hashmi R, Al-Hanawi MK
Husbands' Knowledge of Breast Cancer and Their Wives' Attitudes and Practices Related to Breast Cancer Screening in Saudi Arabia: Cross-sectional Online Survey
J Med Internet Res 2021;23(2):e25404
URL: https://www.jmir.org/2021/2/e25404
doi:10.2196/25404
PMID:33629959

Original Paper

Income-Generating Processes of Free Web-Based Digital Health Tools to Engage Patients: Qualitative Analysis

Claudia Lai[1], BScPharm, MSc, PhD; Raisa Deber[1], PhD; Alejandro R Jadad[1,2], MD, DPhil; Aviv Shachak[1,3], PhD

[1]Institute of Health Policy, Management and Evaluation, Dalla School of Public Health, University of Toronto, Toronto, ON, Canada

[2]Department of Anaesthesiology and Pain Management, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[3]Faculty of Information, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Claudia Lai, BScPharm, MSc, PhD
Institute of Health Policy, Management and Evaluation
Dalla School of Public Health
University of Toronto
155 College Street, Suite 425
Toronto, ON, M5T 3M6
Canada
Phone: 1 416 978 0998
Email: claudia.lai@utoronto.ca

Abstract

Background: In recent years, digital tools have become a viable means for patients to address their health and information needs. Governments and health care organizations are offering digital tools such as self-assessment tools, symptom tracking tools, or chatbots. Other sources of digital tools, such as those offered through patient platforms, are available on the internet free of charge. We define patient platforms as health-specific websites that offer tools to anyone with internet access to engage them in their health care process with peer networks to support their learning. Although numerous social media platforms engage users without up-front charges, patient platforms are specific to health. As little is known about their business model, there is a need to understand what else these platforms are trying to achieve beyond supporting patients so that patients can make informed decisions about the benefits and risks of using the digital tools they offer.

Objective: The aim of this study is to explore what patient platforms are trying to achieve beyond supporting patients and how their digital tools can be used to generate income.

Methods: Textual and visual data collected from a purposeful selection of 11 patient platforms from September 2013 to August 2014 were analyzed using framework analysis. Data were systematically and rigorously coded and categorized according to key issues and themes by following 5 steps: familiarizing, identifying a thematic framework, indexing, charting, and mapping and interpretation. We used open coding to identify additional concepts not captured in the initial thematic framework. This paper reports on emergent findings on the business models of the platforms and their income-generating processes.

Results: Our analysis revealed that in addition to patients, the platforms support other parties with interests in health and information exchanges. Patient platforms did not charge up-front fees but generated income from other sources, such as advertising, sponsorship, marketing (eg, sending information to users on behalf of sponsors or providing means for sponsors to reach patients directly), supporting other portals, and providing research services.

Conclusions: This study reports on the mechanisms by which some patient platforms generate income to support their operations, gain profit, or both. Although income-generating processes exist elsewhere on social media platforms in general, they pose unique challenges in the health context because digital tools engage patients in health and information exchanges. This study highlights the need to minimize the potential for unintended consequences that can pose health risks to patients or can lead to increased health expenses. By understanding other interests that patient platforms support, our findings point to important policy implications, such as whether (and how) authorities might protect users from processes that may not always be in their best interests and can potentially incur costs to the health system.

(J Med Internet Res 2021;23(2):e23654) doi:10.2196/23654

KEYWORDS

digital health; patient engagement; eHealth; health information

Introduction

In recent years, digital tools have become a viable means for patients to address their health and information needs. Governments and health care organizations worldwide are offering digital tools to engage patients, such as self-assessment tools [1], apps to track daily symptoms [2], or chatbots to provide questions and answers [3]. In addition to government and health care organizations, digital tools, such as those found on patient platforms, are now freely available on the internet. Patient platforms are defined as health-specific websites that offer virtual tools to engage patients in their health care process, with peer networks to support user learning (eg, PatientsLikeMe, WebMD). Virtual tools can support the sharing of information that would otherwise be difficult to access, such as personal health information and experiences shared by other patients, thereby enabling some patients to actively learn virtually anywhere at any time, and to receive and provide peer support [4].

Despite the potential benefits, at least for patients who prefer to take an active role in managing their health issues, patient platforms introduce new challenges because their digital tools are not subject to the same rules, regulations, and norms as those offered by health care providers or health care organizations. In contrast, digital tools offered by health care providers or health plans, for example, are governed by privacy legislations such as the *Health Insurance Portability and Accountability Act* in the United States or the *Personal Health Information Protection Act* in Ontario, Canada. In the United States, the US Food and Drug Administration (FDA) is addressing regulation for software as medical devices and regulates a small subset of health apps at this time, for example, those that pose a higher risk if they do not work as intended [5,6]. In other countries such as Canada, new approaches to regulate digital health technologies are not yet in place.

Although the majority of digital tools freely available on the internet from patient platforms are not governed by health authorities, they can influence how patients manage their health issues and whom they consult with for health care advice (eg, health care providers or other patients with related disease or drug use experience). A US survey reported that 1 in 5 users decided to stop medication and 6% to 21% changed their physician as a result of using a patient platform [7]. These actions could influence how patients manage health issues in ways that might benefit or potentially harm them. Two factors that may influence how patient platforms engage their users include stakeholders whose interests the platform serves, including, but not limited to, health care providers, pharmaceutical companies, and insurance companies, and the business model or income-generating mechanisms of the platform. With a growing number of patients using social media tools [4,8], understanding these issues can help patients make informed decisions about the benefits and risks of using patient platforms and inform policy makers on whether and how they might be regulated. Thus, the purpose of this study is to gain insights into what patient platforms are trying to achieve beyond supporting patients, what other stakeholders they support, and how they generate income.

Methods

Overview

A total of 11 patient platforms were purposefully selected, as described later. Textual and visual data collected from these platforms from September 2013 to August 2014 were analyzed using framework analysis [9]. Ethics approval was obtained from the University of Toronto Research Ethics Board.

Platform Selection

Patient platforms were selected from academic literature, patient advocacy websites, industry websites, and internet searches using various search terms such as *social media platforms*, *patient tools*, and *online patient communities*. Platforms were chosen to partly capture variability across the following attributes: (1) for-profit and not-for-profit platforms; (2) platforms supporting single or multiple conditions; and (3) platforms supporting different disease dimensions, including rare conditions (as defined by the platform itself), mental health conditions, and progressive chronic conditions. Progressive chronic conditions were selected from 3 distinct types of illness trajectories, as described by Murray et al [10]: Trajectory 1, which refers to chronic conditions with short periods of evident decline (eg, incurable cancer); Trajectory 2, which refers to chronic conditions with long-term limitations and intermittent episodes of acute deterioration (eg, multiple sclerosis [MS]); and Trajectory 3, referring to chronic conditions with a prolonged gradual decline (eg, neurological conditions) [10]. We deliberately explored platforms offering tools in English (from the United States and Canada), which support patients who can have high but diverse health and information needs. This platform selection strategy allowed for the discovery of concepts and what platforms have in common and unique features in their situated context [11]. Table 1 describes the rationale for case selection using these platform dimensions.

Table 1. Platform dimensions guiding case selection.

Dimension	Platform selection	Rationale
Number of disease groups	• Specific group of patients • Multiple patient groups	Platforms that might achieve higher levels of patient engagement by pooling resources to support a specific group of target users versus platforms that might have rich resources to target a broader range of users
Disease-related characteristics	• Rare disease (eg, amyotrophic lateral sclerosis) • Nonprogressive conditions (eg, anxiety disorder, physical disability, diabetes) • Progressive conditions (eg, Trajectory 1: steady progress and usually a clear terminal phase such as cancers; Trajectory 2: long-term limitations and gradual decline with intermittent episodes of acute deterioration and some recovery such as organ failure or multiple sclerosis; and Trajectory 3: prolonged gradual decline such as neurological conditions)	Platforms that might achieve higher levels of patient engagement by targeting potential users with high health and information needs

We made efforts to select platforms that provide rich information [12], such as platforms with high utilization (eg, number of registered users and user postings, site traffic as measured by industry tools such as Alexa Rank), and operating for long periods (website start date and date of user postings). Platforms with outdated content (ie, no new content in the last 6 months) or limited user postings were excluded from the analysis. This allowed us to explore how they might promote health and encourage information sharing. No additional platforms were added postanalysis because recurring patterns were discovered across the range of platforms [13], reaching saturation on a set of concepts. Table 2 presents the characteristics of the sample based on the aforementioned selection criteria.

Table 2. Characteristics of the selected platforms.

Platform	Single or multiple	Disease dimension	For profit or not for profit	Year started
A1	Single condition	• Rare condition	Not for profit	2006
A2	Multiple conditions	• Multiple related chronic conditions • *Trajectory 3: prolonged dwindling*	Not for profit	1993
A3	Single condition	• Chronic condition • *Trajectory 1: short period of evident decline*	Not for profit	2000
A4	Multiple conditions	• Multiple different conditions	For profit	2005
A5	Multiple conditions	• Multiple different conditions	For profit	1994
A6	Single condition	• Chronic condition • *Trajectory 2: long-term limitations with intermittent episodes*	Not for profit	N/A
A7	Single condition	• Multiple related chronic conditions • *Trajectory 2: short period of evident decline*	For profit	2010
A8	Multiple conditions	• Multiple different conditions	For profit	2004
A9	Multiple conditions	• Multiple different conditions	For profit	N/A
A10	Single condition	• Mental health condition	For profit	N/A
A11	Single condition	• Mental health condition	For profit	2000

Data Collection

Textual and visual data were collected from the platforms' homepages and other webpages relating to the platforms. Because websites are dynamic and subject to changes in design, content, and link structures (both from and to the website), data were captured as screenshots, saved, and stored for offline analysis [14]. These included 302 documents and screenshots collected from September 2013 to August 2014, saved, and stored for offline analysis using the NVivo 11 qualitative data analysis software. Table 3 describes the types of data collected, such as frequently asked questions, *About Us* pages, descriptions of tools, mission statements, registration forms, terms of use, and privacy policies.

These sources offered rich data to analyze what the platforms were trying to achieve, how, and for whom. Only publicly available data that did not require site registration were collected for analysis. No specific platforms, platform users, user groups, or communities are identified in this publication by name. Although platforms may be identifiable, the names of the platforms are masked, as our goal is to illustrate broader issues that can occur on any patient platform and not to single out any individual platform or platform owner. Interested readers may contact the authors for a list of platforms.

Table 3. Data from website home pages and other webpages.

Data collected	Example of data collected	Total number of web pages or documents
Website home pages	Site purpose; platform characteristics (eg, target user groups, number of registered users); description of the platform tools	12
Frequently asked questions	Platform tools, funding sources, and rules pertaining to the website use	32
Site owner information	About us, site, site owners or investors, site mission, testimonials, news releases, and owner webpages	40
Registration process	Registration form	12
Sponsors and partners	Sponsors, partners, and promotional information for the sponsors and partners	37
Tools	Tools and description of tools	133
Terms of use	Policies relating to the use of the platform; how to dispute issues; how to remove user information or terminate account	11
Privacy policies	Policies relating to how user information are collected and used	11
Other rules or policies	Rules relating to forums, advertising and editorial policies, and other rules governing the platform use	5

Data Analysis

Data were systematically tracked and analyzed using framework analysis, which offers a systematic procedure for *sifting, charting, and sorting material according to key issues and themes* [9] by following 5 steps: familiarizing, identifying a thematic framework, indexing, charting, and mapping and interpretation. Preliminary data were systematically and rigorously coded and categorized according to the initial thematic framework, as described elsewhere [15]. Open coding was used to identify additional concepts not captured in the initial thematic framework [16]. Data were subsequently grouped into additional categories, as new concepts were identified from the data. Finally, recurring and overarching themes, unique themes, relationships between various themes and other contextual factors were identified to form theoretical propositions. The coding of data and emergent themes were discussed at bimonthly work sessions with the first and last authors (CL, AS) to reduce potential bias that comes from a single researcher and to increase reliability in observations. Emergent themes were discussed at regular meetings with the research team, including all authors of this paper. This paper reports on concepts from the category of platforms' business model and income-generating processes, which were identified through the aforementioned process of open coding outside of the initial framework.

Results

Platform Sponsors and Partners

Our analysis revealed that in addition to supporting patients, the platforms support other parties with interests in health and information exchanges on the patient platforms. These other parties include patient advocacy organizations, pharmaceutical companies, insurance companies, hospitals and health care providers, employers, health plan administrators, regulators and government organizations, corporations, foundations, and individual donors. Platforms referred to these other parties as *sponsors*, or *partners*, where the distinctions were not always clear; for example, *sponsors* referred to other parties who donated funds to the platform. However, *sponsors* can also provide or fund specific tools or purchase services such as advertising or marketing services. *Partners* referred to other parties, such as patient advocacy organizations or researchers who might collaborate with or without financial arrangements. However, *partners* can also provide tools or purchase services such as clinical trial recruitment services or marketing services. Table 4 provides a summary of sponsors and partners and their relationships with platforms.

Table 4. Partners and sponsors and their relationships with platforms.

Partners and sponsors	Relationship with platforms
Patient advocacy groups	Platforms owned by patient advocacy organizations (A3, A6); facilitated forums for patient advocacy groups (A4, A8) and supported patient advocacy platforms (A11)
Pharmaceutical companies, researchers, and insurance companies (eg, partners or stakeholders who may purchase data or research services)	Pharmaceutical companies, researchers, and other health companies (such as insurance companies) listed as partners or sponsors (A3, A4, A6, A7, A8, A9, A10, and A11)
Hospitals and health care providers (eg, health care providers who may purchase tools for their clients)	Hospitals and health care providers listed as partners and sponsors (A1, A7); hospitals and health care organizations described as clients (A5, A7, A9)
Employers, health plan administrators, regulators, and governments (eg, organizations that may purchase tools for a particular group of users)	Employers, health plan administrators, regulators, and *innovative government organizations* listed as partners or sponsors (A5, A7, A9, A11)
Corporations, foundations, and individual donors	Platforms seeking donations from corporations, foundations, and individual donors (A1, A3, A6)

Income-Generating Mechanisms

All the platforms we studied offered tools free of charge and generated income in other ways, such as through advertisements, fundraising and donations, sponsored tools, marketing services, tools to support other platforms or portals, and research services. Examples are illustrated in Table 5 and described in detail later.

Table 5. Examples of data on income-generating processes.

Item	Category	Examples
a	Advertising	• We also may use Behavioral Advertising cookies which are a way of providing advertisements on the websites you visit and making them more relevant to your interests. (A6)
b	Fundraising	• Shop at AmazonSmile and Amazon will make a donation to [A1]: Get started (A1) • Welcome to Shop for [A3]. Our partners have selected or created items specifically for the benefit of [A3]. A portion of the sale proceeds from your purchase of these items will be donated to [A3], which enables us to continue to provide our online resource to the millions of people we serve each year. Thank you for your contribution. (A3) • Advocacy and Fund-Raising Fundraising Opportunities for [A3] • 16 Topics 252 Posts • Help support our organization, [A3]! Please note, this is NOT a forum for fundraising for other organizations. (A3)
	Donations	• Thanks To Our Sponsors • Want to sponsor this site? Corporate and individual sponsors welcome… Help us help… Help us change even more lives by donating spare change. (A1) • Thank you for your donation! Here is a forum to share with others who you're honoring with your donation. (A3)
c	Sponsored tools	• This dictionary is a compilation of numerous complex breast cancer terms defined in plain English. This program was developed by [A3], and is sponsored in part by an unrestricted educational grant from Bristol-Myers Squibb Oncology, and launched in partnership with... (A3) • Start now: Each module below includes information to help you learn and manage your disease including: Videos, publications, worksheets, links to relevant web pages… [A6] gratefully acknowledges these educational grants to support this project. Bayer HealthCare. Biogen, Novartis, Sanofi Genzyme, Teva Neuroscience (A6)
d	• Marketing • Coupon to save money	• We also provide a voluntary opt-in service to allow partners to directly communicate with patient members through our system. To learn more, see how we make money or read details on what is shared and sold in our Privacy Policy. (A8) • We make managing chronic health conditions easier. We'll work with your doctor and insurance provider to get your medications delivered right to your door at little or no cost. • Save Money Every Month. What if you could save money every single month on the prescription medications you already need and use?...No Insurance? No Problem. What if you could save money every single month on the prescription medications you already need and use? What if you could pay less for your prescribed medications, even without insurance? Believe it or not, such a thing is possible. But don’t take our word for it, get in touch with us and find out for yourself! (A10)
	Recruit patients	• A diabetes pharma brand sought to increase awareness and patient enrollment through branded placement within targeted content on [a diabetes patient platform]. Their goal was to recruit patients at a cost of under $100 per patient enrollment. • A blood pressure pharma brand presented a coupon to a narrow, qualified audience using our xxx program. Of more than 28k leads delivered, nearly 10% converted to a prescription, well over target • A pharmaceutical diagnostics brand sought qualified leads through our xxx program. We delivered more than 300k qualified leads and 27k new prescriptions. The cost per new patient was $60 and offered a value of $280 to the brand. They quadrupled their [platform] lead budget the following year. (A10 site owner)
e	Support other portals	• Important note for clinic patients. • If you have been directed to [A7] by your clinic, do not join here. Instead, use the link that was sent to you via email, or sign up through your clinic's website. (A7)
f	Research	• [A8] may also periodically ask Members to complete short surveys about their experiences (including questions about products and services). Survey responses are analyzed, combined with Members’ Shared Data and shared with and/or sold to Partners. Member participation in these surveys is not required, and refusal to do so will not impact a Member’s experience on the Site. (A8) • We take the information patients like you share about your experience with the disease and sell it to our partners (ie, companies that are developing or selling products to patients). These products may include drugs, devices, equipment, insurance, and medical services. Except for the restricted personal information you entered when registering for the site, you should expect that every piece of information you submit (even if it is not currently displayed) may be shared with our partners and any member of [A8], including other patients. (A8)

Advertisements and Sponsored Content

Most patient platforms, including both for profit and not for profit, generated income by posting advertisements and sponsored content at prominent places on their website. Posting advertisements was framed as beneficial for users, such as "to provide you with our award-winning content at no cost to you" (A9), for the "convenience to patients" (A5), or for helping patients "make educated healthcare choices" (A10). Platforms also noted how they tailored advertisements delivered to their users (examples on behavior advertising are given in Table 5). Advertisements were typically labeled as such. For example, 1 platform noted how they took "steps to ensure that you can clearly identify content that is provided by and is under the editorial control of our sponsors before you view it, so you can make an informed decision as to whether or not to view it" (A9). However, this was not always the case. Platforms sometimes did not make a clear distinction between content intended for the benefit of patients (eg, evidence-based health information) and content intended to benefit sponsors (eg, advertisements or sponsored content). For example, the terms of use for platform A7 described "marketing material from pharmaceutical manufacturers and company information and data about cancer care" in the same sentence as *articles*, *news reports*, or *calculation tools* intended to educate patients. In another case, platform A10 marketed its editorial page for posting sponsored content, as noted below:

> *Sponsorship*
>
> *Highly marketable full editorial page –your content or ours.* [A10 owner site]
>
> *LEVERAGE OUR BRAND TRUST*
>
> *Let us introduce you to our users with a custom performance-based campaign.* [A10 owner site]

Fundraising, Individual, and Corporate Donations

Three platforms (A1, A3, and A6), which operated not for profit, sought financial support through fundraisers, individual donations, corporate sponsorships, and grants. These platforms were condition-specific platforms and sought funds to support a rare disease (A1), cancer (A3), or chronic condition (A6). For example, we found instances where platforms asked users to shop at designated retailers from which the platform receives rewards for purchases (A1), buy mugs or products of which a portion of the sales are donated to the platform (A3), or participate in forums dedicated to fundraising opportunities (A3), as illustrated in Table 5. Donations were also sought from individual donors or corporate sponsors, where sponsors are acknowledged on the platform. We also found tools created from sponsorships or grants, which are described later.

Sponsored Tools or Disease Awareness

Five platforms (A1, A3, A4, A9, and A10), including both for-profit and not-for-profit platforms, provided users with sponsored tools, referring to tools produced or funded by platform sponsors or partners, such as pharmaceutical companies. Sponsored tools included tools peer reviewed by clinicians, tools produced through *educational grants*, or tools for disease awareness, as illustrated in Table 5, sponsored tools. The following is an example of a sponsored self-assessment tool funded by 2 pharmaceutical companies that sell drugs to treat MS:

> *Multiple Sclerosis Assessment*
>
> *This content is selected and controlled by [A9's] editorial staff and is brought to you by EMD Serono, Inc. and Pfizer Inc.*
>
> *How Well Are You Managing MS?*
>
> *… Answer a few questions, and you'll get:*
>
> *Treatment options for your type of MS*
>
> *Information about the progression of the disease*
>
> *Tips for dealing with symptoms while still enjoying life* [Reviewed by [name]MD on [A9] (A9)]

Sponsored tools were presented as being mutually beneficial for sponsors who might also be interested in raising disease awareness.

Marketing Services

Seven platforms (A3, A4, A6, A7, A8, A9, and A10), including both for-profit and not-for-profit platforms, generated income by providing marketing services (Table 5). Users were often asked to consent to receiving marketing material to "inform you of other offers, services, or websites available from [A10] or third parties including our advertising partners" (A10). In some cases, the platforms provided sponsors with a means "to directly communicate with patient members through our system" (A8). Users were sometimes required to provide implicit consent for the sharing of their personal information with site sponsors during the site registration process or when they registered for using specific platform tools, such as patient communities, as noted below:

> *When you register to join a [A10] community and/or register for offers available through our advertising partners, you consent to sharing information about yourself, Personally Identifiable Information, so that we can make our services and the services of our partners available to you.* [A10]

In addition, some platforms generate income by recruiting patients, such as for clinical trials. For example, platform A4 described how they help patients "find relevant clinical trials by inviting them to connect with researchers seeking qualified participant[s]," where the "clinical trial sponsors pay" platform A4 for the service. Besides clinical trial recruitment, 1 platform generated income by providing coupons to users, which was marketed to help them "save money on their healthcare costs" (A10). On the platform owner's website, their services were marketed to sponsors for recruiting patients to change medical interventions to those of the platform's sponsors (eg, blood pressure and diabetes medications or diagnostic products), as illustrated in Table 5.

Tools to Support Health Portals or Other Platforms

Three platforms (A7, A9, and A11), which operated for-profit platforms, marketed their tools (or services) to support the health portals of other organizations. For example, platform A7 advised "clinic patients" to sign up, or login using "the link that was sent to you via email or sign up through your clinic's website,"

as noted in Table 5. Platform A9 described offering their external users with additional services, such as "online and offline health risk assessments, lifestyle education and telephonic health coaching" (A9). Similarly, platform A11 described how their publicly available patient platform offered tools to engage users free of charge for the purpose of *beta-testing software* (eg, testing software updates or new tools), which can later be marketed to other organizations, such as not-for-profit organizations, or pharmaceutical companies.

Research Services

Two platforms (A8 and A9), which operated for-profit platforms, generated income by supporting research services in 1 of the 2 ways. First, platforms described how they conducted surveys or questionnaires for other parties, as illustrated in Table 5. Beyond surveys or questionnaires, 1 platform (A8) described how they are available to support the writing of grant proposals or protocols for review boards, as illustrated below:

> *We are proud to collaborate with some of the leading research institutions in the world on useful and interesting academic research. Please write to the research team with your initial research proposal. If we think a research project has the potential to benefit our users, we would be happy to assist you in writing a grant proposal and helping to describe what we do for your local Internal Review Board (IRB). The proportion of funding we would receive depends on a number of factors including the contribution of our staff to the design, the difficulty of accessing the specific population of interest, and the source of funding.* [A8]

Second, platforms collected a repository of user-generated health data through the platform itself, such as through user profiles or self-tracking tools, as illustrated below:

> *Track your healthcare*
>
> *Chart your health over time and contribute to research that can advance medicine for all.* [A8]
>
> *How is the optional background information used? The more other users know about the users rating the drugs, the more valuable those ratings become. An 18-year-old college student might have a different experience with a particular medication than a post-menopausal woman of 60 on the same drug. Once we have a good sampling of this data, we'll begin letting you search the effectiveness of a drug in individuals similar to you.* [A9]

Although the repository of data can support other users doing research for themselves, 1 platform (A8) was transparent in disclosing their intention to sell the information to "companies that are developing or selling products to patients" which included companies that sell "drugs, devices, equipment, insurance, and medical service" (A8), as noted in Table 5. The information users were required to share included their health profile, biographical information, condition or disease information, treatment information, symptoms, health scores or charts over time, laboratory results, and *connections with other people* on the platform. In fact, this platform advised users to expect *almost any piece of information they submitted*, including information which may not be accessible to the users themselves (or *not currently displayed*), to be shared or sold (except restricted data such as their email address, name, and physical address). This platform (A8) was the only one transparent about this and had even warned users of the potential risks of sharing information online:

> *When sharing information online about your health or a specific condition, you should know there is always a risk that someone could use this information against you. For example, medical and life insurance companies have clauses that exclude pre-existing conditions or employers may not want to employ someone with a high-cost or high-risk disease. We know these risks are real.* [A8]

Discussion

Overview

Despite the introduction of new ways for patients to participate in their health care process and the growing availability of institutional patient-oriented health information systems such as patient portals [17-19], patients often look to digital tools freely available on the internet to address their unmet health and information needs. This study identified several ways by which both for-profit and not-for-profit platforms generated income. These income-generating processes can have important implications for the health of patients and incur costs to the health care system, which will be discussed in detail later.

Advertising

Besides health information, patient platforms can act as a channel to deliver advertisements directly to patients. Advertisements were justified by platforms as necessary for the provision of digital tools to patients without charging up-front fees. Although platforms might claim that users have the choice to ignore advertisements, the labeling of advertisements does not necessarily negate the risks of increased costs. In previous studies, advertisements sent directly to patients, referred to as *direct-to-consumer advertising*, have been associated with increased health care costs because pharmaceutical companies tend to promote more expensive drug therapies (eg, new drugs, new formulations of existing drugs, or novel devices) [20,21]. This may also lead to an increased number of doctor visits or diagnostic testing. For instance, Layton et al [22] illustrated how increased television exposure to testosterone advertisements was associated with the ordering of more testosterone tests and prescribing of more testosterone without testing. Advertising through patient platforms poses unique challenges for many countries that are different from other, more traditional, media outlets. Advertisements posted on the internet from countries with less stringent advertising policies, such as the United States and New Zealand, can reach patients in countries where direct-to-consumer advertising is prohibited [23]. This study adds to the body of literature calling for changes to how advertising and marketing of drugs are regulated globally, as digital tools freely available on the internet enable new channels for advertisements to reach patients across geographic boundaries [24,25].

Sponsored Tools

Various funding arrangements, such as educational grants or sponsorship, have been described as a beneficial way for patients to access tools without paying up-front fees. Whether these tools may be beneficial to patients depends on a number of factors, including the quality of information delivered. Future research is needed to better understand the role that commercial influences might play in the accuracy of the health information provided, which can add to the body of literature on the quality of consumer health information on the internet [26,27]. This study illustrates the possibility of posting editorial content on patient platforms written to support platform sponsors. Moreover, acknowledging corporate sponsorships could encourage individuals to act more favorably toward sponsors in future purchases of products or services. For instance, Jong et al [28] studied the impact of a national campaign to raise awareness of onychomycosis (toenail fungal infection) in Denmark. Despite the availability of various antifungal treatment options, the campaign was followed by an increase in prescriptions for the drug sold by the campaign sponsor only but not for other antifungal drugs for onychomycosis.

In addition, this study identified tools offered by not-for-profit platforms in partnership with sponsors or partners, which may be motivated to raise awareness for rare diseases or specific types of cancers. Although some patients could potentially benefit from earlier detection or treatment of a rare disease or cancer, the full impact of disease awareness tools needs further studies. For example, Mailankody and Prasad [29] raised concerns over a disease awareness campaign for a rare disease (polycythemia vera) that was sponsored by a pharmaceutical company that sells a drug to treat the rare condition. In their view, increased awareness of the rare condition and the drug therapy to treat it could increase the number of patients diagnosed with the condition. However, the diagnostic criteria for the rare disease have not yet been clearly defined. Thus, the authors expressed concerns over the potential misuse of the drug beyond the indications studied and approved by the FDA. They argued that such efforts could drive “wasteful diagnostic testing, overdiagnosis and inappropriate therapy” [29].

As with disease awareness, we identified diagnostic or self-assessment tools provided directly to patients, including tools sponsored by pharmaceutical companies that sell drugs for treating the conditions. Although some patients could potentially benefit from earlier diagnosis, the impact of patients seeking drug therapies based on self-assessments is not known and could potentially cause harm [30-32]. For example, people can ask their physicians to prescribe drug therapies for conditions determined by self-assessments, attempt to obtain these medications directly from internet-based drug distributors (operating from global jurisdictions) [33,34], or order diagnostic tests online [35]. Therefore, this study highlights the need for more research to gain a better understanding of the benefits and risks and patient behaviors associated with the use of sponsored tools, including self-ordering of diagnostic tests and/or drug therapies.

New Ways to Reach Patients

This study identified how marketing services offered by patient platforms could potentially influence the way patients manage their health issues, including the choice of drug therapies or diagnostic tests or enrollment in clinical trials. For instance, coupons offered to help users save money can potentially lead to changes in drug therapies. Mintzes et al [36] raised concerns over pharmaceutical marketing efforts to encourage patients to shift drug use to marketed drug therapies, which may be more expensive. We also found that some platforms connect patients directly with pharmaceutical companies for clinical trials. However, the impact of such direct recruitment is largely unknown, as it deviates from the traditional process where patients consult with their health care provider for optimal therapy. Although some patients could potentially benefit from receiving information from site sponsors (such as pharmaceutical companies), this study highlights the need for more research to better understand the impact of processes that might eliminate traditional intermediaries, such as health care providers who help patients interpret technical information and make informed decisions.

Sharing User Information

This study found numerous examples where user information can be shared with sponsors, which operate for-profit platforms. In one case, the platform was transparent in warning users that their information could be sold to companies (eg, insurance companies) and the associated potential risks. Other platforms were not as transparent. As digital tools, including patient platforms, apps on mobile devices, and wearables, play an increasingly important role in defining what health data are being collected, accessed, and subsequently used for improving health outcomes, there is a critical need for more research and governance in this evolving area.

Conclusions and Policy Implications

In summary, this study revealed many mechanisms by which patient platforms generate income to support their operations, gain profit, or both. Although these income-generating processes can occur on social media in general, patient platforms are health-specific. Thus, they can have health implications on patients and financial impacts on patients and the health care system. Although digital tools may be mutually beneficial to patients, platform owners or operators, sponsors, and partners (eg, enrollment in clinical trials or services that would not otherwise be available to patients), in other cases, they may have negative consequences and potentially harm patients. Given that direct access to platform tools may serve as a way for many patients to gain control over their health issues, there is a need to improve how patients are informed about the risks and benefits of using digital tools freely available on the internet.

Besides transparency in disclosing income-generating processes, there is a broader question as to whether all patients have the capacity to understand the information provided. For instance, it has been reported that patients often lack *the capacity to obtain, understand, and act upon health information and services to make appropriate health decisions on their own* [37] and that improving health literacy is associated with improving

patient engagement [38]. Thus, broader system changes may be needed to increase health literacy at the population level [39,40]. Nevertheless, to ensure that patients can more thoroughly understand and make informed choices about the information and behaviors that these health-specific websites are marketing to them, there is a need for health authorities to expand their role to the evolving digital health landscape [41]. As new digital technologies change how we access and use health data, this study evokes the policy issue of whether and how authorities should protect users from processes that may not be in their best interests.

Limitations and Future Research

Although the collection of publicly available data from patient platforms provided a rich source of information about what platforms were doing, there are some limitations to this study and opportunities for future research. First, we analyzed platforms only but not the behavior of platform users. Future research is needed to explore how people actually use different types of digital tools and their impact on health utilization and expenditures. Second, although the platform selection process was designed to capture the variability of platforms across several attributes, our findings are not representative of all patient platforms, particularly because we only explored English websites (in the United States and Canada). Future studies can explore platforms in other languages or target other patient-important outcomes or dimensions, for example, different cultural values or norms, conditions with high cost, or lack of effective treatment options, which might offer other opportunities for income-generating processes. Third, our findings reflect processes occurring on the selected platforms during the data collection period only (September 2013 to August 2014). We acknowledge that the digital world has evolved since the time of our data collection and analysis. Further, we notice that many general social media platforms have updated their terms of use and privacy policy in light of recent events such as the Facebook or Cambridge Analytica scandal [42]. Similar changes may have occurred in patient platforms' policies. Nevertheless, this study captures a snapshot of the platforms at the time of data collection, which offers a useful way to analyze what can happen. Our findings highlight critical policy implications for patients, providers, and policy makers that continue to be important in our digital health world. Finally, we did not assess the quality of information provided by the digital tools, which could affect whether their business models might offer a win-win situation. Given the current infodemia (or epidemiology of misinformation), where *accurate and timely knowledge translation* can be distorted by various factors such as political and commercial influences [43], and that *fake news* may be distributed or even generated (eg, using artificial intelligence) [42], future research to assess the quality of information on patient platforms will be beneficial. Related to this point, we described our analysis that focused on the written content on business models. Future research can expand our understanding of both the income-generating processes and transparency of disclaimers by exploring the visual presentation of content and the prominence of the placement of the messages on the site. Nevertheless, our findings point to important policy implications in the growing area of online resources provided to patients directly, particularly because platforms are not currently governed to ensure the sharing of accurate information or information that does not do harm.

Conflicts of Interest

None declared.

References

1. COVID-19 self-assessment tool. Ontario Ministry of Health. 2020. URL: https://covid-19.ontario.ca/self-assessment [accessed 2020-05-14]
2. Canada COVID-19 App and Daily Symptom Tracker. Government of Canada. 2020. URL: https://ca.thrive.health/ [accessed 2020-05-14]
3. COVID-19. City of Toronto. 2020. URL: https://www.toronto.ca/home/covid-19/ [accessed 2020-05-14]
4. Singh K, Meyer SR, Westfall JM. Consumer-Facing Data, Information, And Tools: Self-Management Of Health In The Digital Age. Health Aff (Millwood) 2019 Mar;38(3):352-358. [doi: 10.1377/hlthaff.2018.05404] [Medline: 30830832]
5. Digital Health Center of Excellence. US Food and Drug Administration. 2021. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence [accessed 2021-01-13]
6. 21st Century Cures Act. US Food & Drug Administration. 2020. URL: https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/21st-century-cures-act [accessed 2021-01-09]
7. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
8. The Social Life of Health Information. Pew Research Center. 2009 Jun 11. URL: https://www.pewresearch.org/internet/2009/06/11/the-social-life-of-health-information/ [accessed 2021-01-13]
9. Richie J, Spencer L. In: Huberman AM, Miles MB, editors. The Qualitative Researcher's Companion. Thousand Oaks, CA: SAGE Journals; 2002:305-329.
10. Murray SA, Kendall M, Boyd K, Sheikh A. Illness trajectories and palliative care. Br Med J 2005 Apr 30;330(7498):1007-1011 [FREE Full text] [doi: 10.1136/bmj.330.7498.1007] [Medline: 15860828]
11. Denzin NK. Strategies of Multiple Triangulation. In: The Research Act: A Theoretical Introduction to Sociological Methods. New Brunswick, USA: AldineTransaction; 2009:297-313.

12. Patton M. Qualitative research and evaluation methods (3rd ed). In: Integrating Theory and Practice. Thousand Oaks, CA: SAGE Publications; 2002:44-45.
13. Miles MB, Huberman AM. Qualitative data analysis: An expanded sourcebook (2nd Ed). Thousand Oaks, CA: Sage Publications; 1994.
14. Bar-Ilan J, Peritz BC. Evolution, continuity, and disappearance of documents on a specific topic on the Web: A longitudinal study of “informetrics”. J. Am. Soc. Inf. Sci 2004 Sep;55(11):980-990. [doi: 10.1002/asi.20049]
15. Lai C, Jadad AJ, Deber R, Shachak A. Tools for Engaging Patients on Patient Platforms: A Classification Framework. Stud Health Technol Inform 2019;257:256-260. [doi: 10.3233/978-1-61499-951-5-256] [Medline: 30741205]
16. Glaser BG, Strauss AL. The Discovery of Grounded Theory. In: The Discovery of Grounded Theory. Chicago, IL: Aldine; 1967.
17. Deshpande A, Jadad AR. Web 2.0: Could it help move the health system into the 21st century? The Journal of Men's Health & Gender 2006 Dec;3(4):332-336. [doi: 10.1016/j.jmhg.2006.09.004]
18. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008 Aug 25;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
19. Social Media Fact Sheet. Pew Research Center. 2019 Jun 12. URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2021-01-09]
20. Gagnon M, Lexchin J. The cost of pushing pills: a new estimate of pharmaceutical promotion expenditures in the United States. PLoS Med 2008 Jan 03;5(1):e1 [FREE Full text] [doi: 10.1371/journal.pmed.0050001] [Medline: 18177202]
21. Wilkes MS, Bell RA, Kravitz RL. Direct-to-consumer prescription drug advertising: trends, impact, and implications. Health Aff (Millwood) 2000;19(2):110-128. [doi: 10.1377/hlthaff.19.2.110] [Medline: 10718026]
22. Layton JB, Kim Y, Alexander GC, Emery SL. Association Between Direct-to-Consumer Advertising and Testosterone Testing and Initiation in the United States, 2009-2013. JAMA 2017 Mar 21;317(11):1159-1166 [FREE Full text] [doi: 10.1001/jama.2016.21041] [Medline: 28324090]
23. Hébert PC. Direct-to-consumer advertising: end the compromise. CMAJ 2008 Jul 15;179(2):119 [FREE Full text] [doi: 10.1503/cmaj.080927] [Medline: 18625974]
24. Mackey TK, Cuomo RE, Liang BA. The rise of digital direct-to-consumer advertising?: Comparison of direct-to-consumer advertising expenditure trends from publicly available data sources and global policy implications. BMC Health Serv Res 2015 Jun 19;15(1):236 [FREE Full text] [doi: 10.1186/s12913-015-0885-1] [Medline: 26084705]
25. Rozenblum R, Bates DW. Patient-centred healthcare, social media and the internet: the perfect storm? BMJ Qual Saf 2013 Mar 01;22(3):183-186. [doi: 10.1136/bmjqs-2012-001744] [Medline: 23378660]
26. Risk A, Dzenowagis J. Review of internet health information quality initiatives. J Med Internet Res 2001 Dec 26;3(4):E28 [FREE Full text] [doi: 10.2196/jmir.3.4.e28] [Medline: 11772543]
27. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002 May 22;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
28. Jong GW, Stricker BHC, Sturkenboom MCJM. Marketing in the lay media and prescriptions of terbinafine in primary care: Dutch cohort study. BMJ 2004 Feb 26;328(7445):931. [doi: 10.1136/bmj.38007.711481.f7]
29. Mailankody S, Prasad V. Pharmaceutical Marketing for Rare Diseases: Regulating Drug Company Promotion in an Era of Unprecedented Advertisement. JAMA 2017 Jun 27;317(24):2479-2480. [doi: 10.1001/jama.2017.5784] [Medline: 28520871]
30. Ryan A, Wilson S. Internet healthcare: do self-diagnosis sites do more harm than good? Expert Opin Drug Saf 2008 May;7(3):227-229. [doi: 10.1517/14740338.7.3.227] [Medline: 18462181]
31. Fergus TA, Spada MM. Cyberchondria: Examining relations with problematic Internet use and metacognitive beliefs. Clin Psychol Psychother 2017 Nov 16;24(6):1322-1330. [doi: 10.1002/cpp.2102] [Medline: 28621035]
32. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: Systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
33. Gernburd P, Jadad AR. Will spam overwhelm our defenses? Evaluating offerings for drugs and natural health products. PLoS Med 2007 Sep 18;4(9):e274 [FREE Full text] [doi: 10.1371/journal.pmed.0040274] [Medline: 17880258]
34. Ivanitskaya L, Brookins-Fisher J, O Boyle I, Vibbert D, Erofeev D, Fulton L. Dirt cheap and without prescription: how susceptible are young US consumers to purchasing drugs from rogue internet pharmacies? J Med Internet Res 2010 Apr 26;12(2):e11 [FREE Full text] [doi: 10.2196/jmir.1520] [Medline: 20439253]
35. Rockwell KL. Direct-to-Consumer Medical Testing in the Era of Value-Based Care. JAMA 2017 Jun 27;317(24):2485-2486. [doi: 10.1001/jama.2017.5929] [Medline: 28542699]
36. Mintzes B, Barer ML, Kravitz RL, Bassett K, Lexchin J, Kazanjian A, et al. How does direct-to-consumer advertising (DTCA) affect prescribing? A survey in primary care environments with and without legal DTCA. Can Med Assoc J 2003 Sep 2;169(5):405-412 [FREE Full text] [Medline: 12952801]
37. Health Literacy in Canada: Initial Results from the International Adult Literacy Skills Survey. Canadian Council on Learning. Ottawa; 2007. URL: http://www.ccl-cca.ca/pdfs/HealthLiteracy/HealthLiteracyinCanada.pdf [accessed 2021-01-09]

38. McCormack L, Thomas V, Lewis MA, Rudd R. Improving low health literacy and patient engagement: A social ecological approach. Patient Educ Couns 2017 Jan;100(1):8-13. [doi: 10.1016/j.pec.2016.07.007] [Medline: 27475265]
39. Health Council of Canada. Turning what we know into action: A commentary on the National Symposium on Patient Engagement. In: Health Council of Canada. Toronto: Health Council of Canada; Feb 2012.
40. Koh HK, Brach C, Harris LM, Parchman ML. A proposed 'health literate care model' would constitute a systems approach to improving patients' engagement in care. Health Aff (Millwood) 2013 Feb;32(2):357-367 [FREE Full text] [doi: 10.1377/hlthaff.2012.1205] [Medline: 23381529]
41. Liang BA, Mackey T. Direct-to-consumer advertising with interactive internet media: global regulation and public health issues. JAMA 2011 Feb 23;305(8):824-825. [doi: 10.1001/jama.2011.203] [Medline: 21343583]
42. Isaak J, Hanna MJ. User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. Computer 2018 Aug;51(8):56-59. [doi: 10.1109/mc.2018.3191268]
43. Eysenbach G. How to Fight an Infodemic: The Four Pillars of Infodemic Management. In: J Med Internet Res. Canada: JMIR Publications; Jun 29, 2020:e21820.

Abbreviations

MS: multiple sclerosis
FDA: Food and Drug Administration

Edited by G Eysenbach; submitted 20.08.20; peer-reviewed by D Wang, N Ratanawongsa; comments to author 11.09.20; revised version received 08.10.20; accepted 18.12.20; published 03.02.21.

Please cite as:
Lai C, Deber R, Jadad AR, Shachak A
Income-Generating Processes of Free Web-Based Digital Health Tools to Engage Patients: Qualitative Analysis
J Med Internet Res 2021;23(2):e23654
URL: http://www.jmir.org/2021/2/e23654/
doi:10.2196/23654
PMID:33533722

Original Paper

Using the Patient Portal Sexual Health Instrument in Surveys and Patient Questionnaires Among Sexual Minority Men in the United States: Cross-sectional Psychometric Validation Study

Kevon-Mark P Jackman[1], DrPH, MPH; Jeremy Kane[2], PhD, MPH; Hadi Kharrazi[3], PhD, MD; Renee M Johnson[1], PhD, MPH; Carl Latkin[4], MS, PhD

[1]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Department of Epidemiology, Columbia University Mailman School of Public Health, NY, NY, United States

[3]Department of Health Policy and Management, Center for Population Health IT, Johns Hopkins Bloomberg School of Public Health, Batimore, MD, United States

[4]Department of Health, Behavior, and Society, Johns Hopkins Blooomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:
Kevon-Mark P Jackman, DrPH, MPH
Department of Mental Health
Johns Hopkins Bloomberg School of Public Health
624 N Broadway
Hampton House 8th Floor
Baltimore, MD, 21205
United States
Phone: 1 410 955 3910
Email: kjackma2@jhmi.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/3/e28358/

Abstract

Background: Patient portal modules, including electronic personal health records, health education, and prescription refill ordering, may be leveraged to address the sexually transmitted infection (STI) burden, including HIV, among gay, bisexual, and other sexual minority men (SMM). Theoretical frameworks in the implementation sciences highlight examining constructs of innovation attributes and performance expectations as key determinants of behavioral intentions and the use of new web-based health technologies. However, behavioral intentions to use patient portals for HIV and other STI prevention and care among SMM is understudied.

Objective: The aim of this study is to develop a brief instrument for measuring attitudes focused on using patient portals for STI prevention and care among a nationwide sample of SMM.

Methods: A total of 12 items of the American Men's Internet Survey-Patient Portal Sexual Health Instrument (AMIS-PPSHI) were adapted from a previous study. Psychometric analyses of the AMIS-PPSHI items were conducted among a randomized subset of 2018 AMIS participants reporting web-based access to their health records (N=1375). Parallel analysis and inspection of eigenvalues in a principal component analysis (PCA) informed factor retention in exploratory factor analysis (EFA). After EFA, Cronbach α was used to examine the internal consistency of the scale and its subscales. Confirmatory factor analysis (CFA) was used to assess the goodness of fit of the final factor structure. We calculated the total AMIS-PPSHI scale scores for comparisons within group categories, including age, STI diagnosis history, recency of testing, serious mental illness, and anticipated health care stigma.

Results: The AMIS-PPSHI scale resulting from EFA consisted of 12 items and had good internal consistency (α=.84). The EFA suggested 3 subscales: sexual health engagement and awareness (α=.87), enhancing dyadic communication (α=.87), and managing sexual health care (α=.79). CFA demonstrated good fit in the 3-factor PPSHI structure: root mean square error of approximation=0.061, comparative fit index=0.964, Tucker-Lewis index=0.953, and standardized root mean square residual=0.041. The most notable differences were lower scores on the enhanced dyadic communication subscale among people living with HIV.

Conclusions: PPSHI is a brief instrument with strong psychometric properties that may be adapted for use in large surveys and patient questionnaires in other settings. Scores demonstrate that patient portals are favorable web-based solutions to deliver health services focused on STI prevention and care among SMM in the United States. More attention is needed to address the privacy implications of interpersonal use of patient portals outside of traditional health settings among persons with HIV.

(J Med Internet Res 2021;23(2):e18750) doi:10.2196/18750

KEYWORDS

health information technology; sexual behavior; HIV; STI; patient portals

Introduction

Psychometrics in Health Technology Behavior Research

The use of digital strategies to address public health priorities, such as HIV, has increased tremendously in the United States in the last decade [1-4]. To inform implementation strategies, many survey instruments have been developed to measure contextual attitudes about health technology use across patient populations. The National Cancer Institute's Health Information National Trends Survey (HINTS) is a biennial nationally representative survey to assess the impact of the health information environment. It includes items useful to describe individual perceptions about the privacy and security of electronic medical records in a national sample [5]. Instrument development has been applied to assess attitudes about health technology use in specific populations, such as among older adults [6]. Psychometric constructs assessed within instruments are sometimes the basis of path models developed to measure behavioral intentions among consumers using health information technology, such as the Health Information Technology Acceptance Model [7]. This study focuses on developing an instrument for a neglected area of study, patient portal use for the prevention and care of sexually transmitted infections (STIs), including HIV.

Patient Portals

Patient portals provide patients with secure web-based access to health information, such as laboratory results and prescription medications. The web-based personal health information that a consumer accesses when using their patient portal is a type of health information technology (Health IT), referred to as personal health records (PHRs) [8,9]. Health IT is a broad concept that encompasses an array of technologies applied in health systems, such as PHRs, electronic health records (EHRs), or e-prescribing [10]. Logging on to a patient portal, users may message doctors, locate health education content, and a variety of options, depending on the Health IT platform. Certification for Health IT products is established by standards, implementation specifications, and certification criteria adopted by the US Secretary of Health and Human Services [11]. Data security and privacy measures, such as encrypting authentication credentials, are outlined within standards [12]. This is important because privacy concerns and mistrust are known barriers to PHR adoption [13,14]. Nevertheless, among US sexual minority men (SMM), patient portals are highly acceptable for delivery of comprehensive sexual health services that support HIV preventative behaviors, such as disclosing STI PHRs with sex partners [15,16].

In a 2016 market of approximately 186 certified vendors, 92% of US hospitals contracted either Cerner, MEDITECH, Epic Systems, CPSI, McKesson, and MEDHOST to supply certified Health IT [17]. To be certified, EHR vendors must include a patient portal module and use Fast Healthcare Interoperability Resources specifications to allow communication across vendor platforms [12,18]. Examples of patient portals are Epic's MyChart, Veterans Administration's My HealtheVet, Geisinger's MyGeisinger, and Kareo's Kareo EHR Patient Portal. There are over 300 unique Health IT products with 2015 certification that allow patients (ie, consumers) to view, transmit, and download patient medical data to a third party [19]. PHR systems may also be standalone and untethered to a health system's EHR system [9].

As a result of federal incentive programs, namely, the Merit-Based Incentive Payment System programs Advancing Care Information, which have replaced Meaningful Use, patient portal access in health care settings has grown exponentially in the last decade and provides an excellent opportunity to deliver important health communications to patients [20-22]. By 2015, 95% of hospitals and 63% of office-based physicians had IT solutions allowing patients to view web-based health records [22,23]. Patient medical records are often perceived as highly private and only to be viewed by the patient and authorized representatives, which are based on important ethical and societal concerns of privacy [24-26]. However, portals also allow patients to interpersonally use information in PHRs outside of health care settings, such as sharing laboratory results or prescriptions with family, friends, and partners when promoting health behaviors [27]. Furthermore, few empirical studies have measured the preferences for and potential health benefits of the patient portal for sexual health and well-being on both individual and interpersonal levels.

The Utility of Patient Portals for STIs

To date, little information is available on patients' perceptions of patient portals and their potential use for the prevention and care of STIs, including HIV [28-30]. STIs are substantial public health problems, and they may be particularly well suited for prevention and management with patient portals. With billions of dollars in US health care costs, an estimated 20 million new infections each year, and as major causes of preventable disease, STIs are serious and growing public health threats [31]. In 2018, there were 2,457,118 combined cases of chlamydia, gonorrhea, and syphilis alone in the United States, continuing the trend of record-breaking numbers within the last decade [32]. The increasing prevalence of multidrug-resistant gonorrhea and the rise in syphilis infections also signal the urgent need for more effective prevention strategies [33,34].

Inefficient access and adherence to HIV medication among people living with HIV and pre-exposure prophylaxis (PrEP) among some SMM risk groups contribute to the burden of new HIV infections. In fact, persons with HIV who attain virologic suppression cannot transmit HIV to others, as regular PrEP use can effectively prevent the acquisition of HIV infection [35,36]. Adherence requires compliance with outpatient provider visits, filling in prescriptions, and taking antiretroviral therapy (ART) or PrEP over time; however, nonadherence can increase HIV disease progression and risk of disease transmission. Patient portals may effectively be able to create an easy-to-use system for organizing health information and be helpful in engaging SMM in care on all fronts of HIV infection.

Given the stigma about same-sex sexuality and talking about HIV and STIs, portals can also bridge communication gaps with providers and sex partners. According to the Centers for Disease Control and Prevention, gay, bisexual, and other SMM are disproportionately impacted by syphilis, HIV, and other STIs [37,38]. Factors such as patient-provider communication barriers and anticipated health care stigma experienced among SMM are known to hinder the use of STI preventive services such as testing [39-41]. In addition, controlling HIV is harder among persons with mental health issues, resulting in lower rates of HIV virologic suppression [42,43]. In a call to action for comprehensive HIV services for SMM, Beyrer et al [44] recommend health providers to provide integrated mental health services for SMM. SMM today have demonstrated high acceptability of medical and public health practice supported by mobile devices or mobile health (mHealth) in HIV interventions; however, these solutions have largely operated outside of the Health IT aegis of EHR and integrated PHR systems [45]. Previous studies examining perceptions of patient portal use for HIV care and STI prevention have demonstrated high acceptability and concern about potential breaches of privacy [28,30,46].

Study Purpose

With this changing environment, new studies continue to burgeon modeling the implementation of health innovations for behavioral interventions focused on SMM. However, little data are available on the attributes and expectations of patient portals among SMM broadly across the United States. Theoretical frameworks, such as the Diffusion of Innovation theory and the extended unified theory of adoption and use of technology, are often used in behavioral sciences to inform implementation of new health technologies across consumer populations [6,47,48]. The theories posit that attributes of health technology and performance expectations are drivers of behavioral intentions to use health technology, and further behavioral intentions and habits determine behaviors or patterns of use. At the foundation of optimizing the design and uptake of new health technologies is scaling consumer perceptions about technology. Therefore, the overall goal of this study is to evaluate the psychometrics and adaptation of an instrument for measuring attitudes about using patient portals for STI prevention and care. Within-group comparisons of instrument scale scores are examined for categorical variables, including age, US region, last STI test, STI diagnosis, anticipated health care stigma, and mental health status.

Methods

Study Overview

Data were obtained from the 2018 American Men's Internet Survey (AMIS). AMIS is an annual cross-sectional web-based survey of US residents who are aged at least 15 years; are cisgender male; and are gay or bisexual or have ever had sex with a man [49]. The study was conducted in compliance with federal regulations governing the protection of human subjects, and the study protocol was reviewed and approved by the Institutional Review Board at Emory University.

A randomized subset of 4647 AMIS participants was presented with survey items focusing on using patient portal services for STI prevention and care. Only survey participants reporting the ability to view their web-based health records were included in the analysis. Of these 2566 participants with patient portal access, the participants with *do not know* and *refuse to answer* responses on patient portal items were excluded from the analysis, resulting in a final analytic sample of 1375 men.

Measures and Instrument Development

The Patient Portal Sexual Health Instrument (PPSHI) is a scale designed to evaluate perceptions of using patient portals to promote STI prevention and care behaviors. Items were originally adapted from a study among coed students at a historically Black university characterizing the perceived role of patient portals in supporting STI prevention behaviors—the Electronic Sexual Health Information Notification and Education (eSHINE) Study. In total, the eSHINE Study-Patient Portal Instrument (eSHINE-PPI) consisted of 19 items representing 4 subscales: (1) sexual health engagement (4 items), (2) informational resource compatibility (3 items), (3) valuation of services (5 items), and (4) PHR impact (7 items). A complete list of eSHINE-PPI items, factor loadings corresponding to unique subscales, and reliability coefficients can be found in Multimedia Appendix 1 [29,50]. To explore a more parsimonious instrument, we adapted 12 items from 3 eSHINE-PPI subscales to a nationwide survey of SMM. We focused on eliminating the eSHINE-PPI subscale items with the lowest factor loadings. Responses were also reduced from 7-point Likert scales to 4-point ordinal scales. We then adapted the eSHINE-PPI instrument to the AMIS-2018 survey.

Adapting the Sexual Health Engagement Subscale

The lowest loading item on the sexual health engagement subscale, "I plan to manage my medical records with PHRs in the future," was eliminated. The final adapted subscale (3 items) measured perceived attributes of using (vs not using) patient portals, specifically that (1) it is a more convenient way to manage my sexual health records, (2) it encourages people to be more aware of their sexual health, and (3) it will help people like me make better sexual health decisions. Responses were coded as follows: 0=strongly disagree, 1=disagree, 2=agree, and 3=strongly agree. Possible scores ranged from 0 to 9.

Adapting the PHR Impact Subscale

The following 4 items were eliminated from the PHR impact subscale: (1) *PHRs make it easier for people to routinely have*

check-in conversations with partners about STI prevention, (2) *Partners using PHRs will start talking about STI prevention earlier in a relationship*, (3) *I would have more discussions with partners about STI testing if PHRs were more commonly used*, and (4) *Using PHRs with a partner builds trust*. The final adapted subscale (3 items) measured agreement with beliefs that sharing STI PHRs with partners will (1) improve communication on HIV and other STIs, (2) improve confidence in the testing information a partner share, and (3) improve control over my sexual health and decision making. These responses were coded as follows: 0=definitely not, 1=probably not, 2=probably, and 3=definitely. Possible scores ranged from 0 to 9.

Adapting the Valuation of Services Subscale

The item *In addition to electronic sexually transmitted disease (STD) results, which services are important for PHRs to include: Access to all of your medical records* was eliminated from the valuation of services subscale. Given the increase in telehealth success, the *Services to communicate with your doctor or health professionals* item was modified to specify *video chat services to communicate with health care providers* [51,52]. In response to recent decade increases in the use of interactive games to improve HIV prevention behaviors and the use of home test kits for STI screening, we added 2 items assessing valuation for *games to promote sexual health* and the *ability to order home test kits for HIV and STDs* to the valuation of services subscale [45,53-58]. For the final adapted subscale (6 items), participants were asked to rate the value of 6 patient portal features: (1) games to promote sexual health, (2) ability to order home test kits for STIs, (3) counseling and resources for people with STIs, (4) telehealth services, (5) ability to locate STI testing centers, and (6) tips or tools for managing sexual health. Responses on patient portal functionality items were coded as follows: 0=no value, 1=low value, 2=moderate value, and 4=high value. Possible scores ranged from 0 to 18.

The Informational Resource Compatibility Subscale (3 items) was not adapted to AMIS. Possible scores for the total AMIS-PPSHI ranged from 0 to 36.

Statistical Analysis

First, we conducted a principal component analysis (PCA) of the 12 items to estimate an appropriate number of factors to retain. The retention of factors was determined by a parallel analysis and examining eigenvalues greater than 1.0 [59]. We used the exploratory factor analysis (EFA) to estimate the factor structure and loadings. Item loadings were examined for each factor using cutoff values ≥0.50. Sampling adequacy was indicated by a Kaiser-Meyer-Olkin (KMO) score greater than 0.80 [60].

We estimated the overall scale and subscale internal reliability with Cronbach α for the new AMIS-PPSHI. Confirmatory factor analysis (CFA) was used to evaluate measures of fit, including the Tucker-Lewis index (TLI), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Thresholds for fit statistics were values close to 0.95 for TLI and CFI, values close to 0.08 for SRMR, and values close to 0.06 for RMSEA [61]. For standardized factor loadings, we used a cutoff of 0.50 [62].

We calculated the total AMIS-PPSHI scale scores for comparisons within group categories, including age, STI diagnosis history, recency of testing, serious mental illness, and anticipated health care stigma. The Kessler 6-item (K6) scale for psychological distress was used to determine the presence of serious mental illness (score 13 or greater) [63]. Anticipated health care stigma was defined as whether the participant was ever afraid to seek health care services because of worrying that someone may learn they have sex with men. Two-sample *t* tests were used to compare binary group differences in scale scores using a *P* value of .05 as the criterion for statistical significance. Analysis of variance (ANOVA) was used to compare group differences in scale scores for variables with more than 2 groups, also using a *P* value of .05 as the criterion for statistical significance.

Results

Sample Description

The sample consisted of 1375 US SMM with a median age of 34 years and an IQR of 25 to 50 years. The demographic characteristics of the study sample are presented in Table 1. Table 2 presents the items included in the questionnaire with calculated mean and mode scores. The KMO score (0.8492) established sampling adequacy. We used a Promax rotation because the factor correlations exceeded 0.32 [64]. On the basis of eigenvalues above 1.0 and parallel analysis, we estimated a 3-factor solution for factor analysis. Factor analysis yielded 3 distinct factors using the 0.50 cutoff and no cross-loading above 0.15.

Table 1. Demographic data for the study population, including age, region, and sexually transmitted infection diagnosis history, American Men's Internet Survey 2018 (N=1375).

Category	Participants, n (%)
Age (years)	
15-24	318 (23.13)
25-29	214 (15.56)
30-39	291 (21.16)
≥40	552 (40.15)
US region	
Northeast (eg, New York and Vermont)	221 (16.07)
Midwest (eg, Illinois and Ohio)	314 (22.84)
South (eg, Florida and Alabama)	505 (36.73)
West (eg, California and Oregon)	335 (24.36)
Race or ethnicity	
Black, non-Hispanic	75 (5.51)
Hispanic	181 (13.30)
White, non-Hispanic	1020 (74.94)
Other or multiple race	85 (6.25)
History of STI[a] diagnosis	
No history of HIV or other STI diagnosis	1087 (79.05)
Living with HIV	118 (8.58)
Recent gonorrhea, chlamydia, or syphilis (and not living with HIV)	170 (12.36)
STI test in the 12 months before the study	
No	1128 (82.04)
Yes	247 (17.96)
Willing to access web-based STI test results	
No	14 (1.02)
Yes	1361 (98.98)

[a]STI: sexually transmitted infection.

Table 2. Descriptive statistics of suggested instrument items, American Men's Internet Survey 2018 (N=1375).

Questionnaire item	Mean (SD)	Mode
Patient engagement with sexual health care items		
Item 1. It is a more convenient way to manage my sexual health records.	2.56 (0.59)	3
Item 2. It encourages people to be more aware of their sexual health.	2.52 (0.60)	3
Item 3. It will help people like me make better sexual health decisions.	2.37 (0.71)	2
Dyadic communication items		
Item 4. Improve communication on HIV and other sexually transmitted infections	2.49 (0.65)	3
Item 5. Improve my confidence in the testing information a partner shares with me	2.54 (0.63)	3
Item 6. Improve control over my sexual health and decision making	2.54 (0.64)	3
Conceptual features (functionality)		
Item 7. Tips or tools for managing sexual health	2.43 (0.68)	3
Item 8. Ability to locate STD[a] test centers and services	2.63 (0.64)	3
Item 9. Video chat for communicating with health care providers	2.16 (0.85)	2
Item 10. Counseling and resources for people with STDs	2.54 (0.68)	3
Item 11. Ability to order home test kits for HIV and STDs	2.59 (0.68)	3
Item 12. Games to promote sexual health	1.54 (1.02)	1

[a]STD: sexually transmitted disease.

The AMIS-PPSHI Structural Model

Table 3 displays the 12 items and their respective loadings across the 3 factors. Items 1-3 loaded on a factor named *sexual health engagement and awareness*, reflecting the perceived attributes of using patient portals to engage in sexual health care. Items 4-6 loaded on a factor named *enhancing dyadic communication*, reflecting the perceived attributes of using STI PHRs to share testing history with partners. Items 7-12 loaded on a third factor named *managing sexual health care*, reflecting the desired functionality of patient portals. Inter-factor correlation were 0.42 between the *sexual health engagement and awareness* and *enhancing dyadic communication* factors. *Sexual health engagement and awareness* and *managing sexual health care* had a correlation of 0.29. The estimated correlation between *managing sexual health care* and *enhancing dyadic communication* was 0.42. The CFA results (Table 4) suggested good fit indices for the 3-factor model (RMSEA=0.061, CFI=0.964, TLI=0.953, SRMR=0.041, and coefficient of determination =0.996).

Table 3. Exploratory factor analysis loadings and Cronbach α values, American Men's Internet Survey 2018 (N=1375).

Questionnaire item	Factor loadings[a,b,c]		
	Factor 1. Sexual health engagement and awareness	Factor 2. Enhancing dyadic communication	Factor 3. Managing sexual health care
Item 1. It is a more convenient way to manage my sexual health records	*0.7994*	−0.0268	−0.0061
Item 2. It encourages people to be more aware of their sexual health	*0.8740*	−0.0231	0.0123
Item 3. It will help people like me make better sexual health decisions	*0.7374*	0.1093	0.0224
Item 4. Improve communication on HIV and other sexually transmitted infections	−0.0171	*0. 7773*	0.0465
Item 5. Improve my confidence in the testing information a partner shares with me	0.0012	*0.8213*	−0.0123
Item 6. Improve control over my sexual health and decision making	0.0494	*0.8047*	−0.0226
Item 7. Tips or tools for managing sexual health	0.0806	0.1323	*0.5531*
Item 8. Ability to locate STD[d] test centers and services	−0.0058	−0.0348	*0.6967*
Item 9. Video chat for communicating with health care providers	0.0199	−0.0478	*0.6382*
Item 10. Counseling and resources for people with STDs	0.0208	0.0039	*0.7415*
Item 11. Ability to order home test kits for HIV and STDs	−0.0202	−0.0057	*0.5621*
Item 12. Games to promote sexual health	−0.0743	0.0741	*0.5269*

[a]Factor loadings above 0.5000 are italicized.

[b]Cronbach α for factors: factor 1, sexual health engagement and awareness, α=.8678; factor 2, enhancing dyadic communication, α=.8689; and Factor 3, managing sexual health care, α=.7888.

[c]Interfactor correlations: $r_{factor1,factor2}$=0.42, $r_{factor1,factor3}$=0.29, and $r_{factor2,factor3}$=0.42.

[d]STD: sexually transmitted disease.

Table 4. Confirmatory factor analysis of standardized factor loadings for Patient Portal Sexual Health Instrument, American Men's Internet Survey 2018 (N=1375).

Questionnaire item	Factor loadings		
	Sexual health engagement and awareness	Enhancing dyadic communication	Managing sexual health care
Item 1. It is a more convenient way to manage my sexual health records	0.7929	N/A[a]	N/A
Item 2. It encourages people to be more aware of their sexual health	N/A	N/A	N/A
Item 3. It will help people like me make better sexual health decisions	N/A	N/A	N/A
Item 4. Improve communication on HIV and other sexually transmitted infections	N/A	0.8119	N/A
Item 5. Improve my confidence in the testing information a partner shares with me	N/A	0.8429	N/A
Item 6. Improve control over my sexual health and decision making	N/A	0.8357	N/A
Item 7. Tips or tools for managing sexual health	N/A	N/A	0.6633
Item 8. Ability to locate STD[b] test centers and services	N/A	N/A	0.6840
Item 9. Video chat for communicating with health care providers	N/A	N/A	0.6202
Item 10. Counseling and resources for people with STDs	N/A	N/A	0.7760
Item 11. Ability to order home test kits for HIV and STDs	N/A	N/A	0.5493
Item 12. Games to promote sexual health	N/A	N/A	0.5339

[a]N/A: not applicable.

[b]STD: sexually transmitted disease.

AMIS-PPSHI Scores

Table 5 presents factor scores for AMIS-PPSHI total and Table 6 factor scores for AMIS-PPSHI subscales by group categories using the sum of scores for variables with factor loadings above a cutoff of 0.50 [65]. The mean (M) and standard deviation (SD) AMIS-PPSHI total score was mean 28.94 (SD 5.14). Mean PPSHI subscale scores were as follows: sexual health engagement and awareness mean 7.46 (SD 1.70), enhancing dyadic communication mean 7.57 (SD 1.71), and managing sexual health care mean 13.90 (SD 3.22). By region, AMIS-PPSHI total scores were moderately higher than average in the South and West compared with the Northeast and Midwest regions. Mean AMIS-PPSHI scores decreased with increasing age category, most notably in the *enhancing dyadic communication* scale ($F_{3,1371}$=10.87; *P*<.001). Participants who were tested 12 months before the study had slightly higher mean scores on sexual health engagement and awareness, mean 7.52 (SD 1.66) versus mean 7.18 (SD 1.83), and enhancing dyadic communication, mean 7.62 (SD 1.69) versus mean 7.38 (SD 1.82). The largest difference in the AMIS-PPSHI score were in comparisons of participants according to their history of HIV or recent STI. The overall highest scores were among people without HIV and without recent STI. Participants living with HIV have the lowest overall AMIS-PPSHI mean scores, mean 27.57 (SD 5.89), primarily because of enhancing dyadic communication scores, mean 6.85 (SD 2.13). There are no significant differences in scores by anticipated health care stigma nor serious mental illness. However, scores were marginally higher among participants with a Kessler 6-item psychological distress scale (K6) score≥13.

Table 5. Psychometrics of the Patient Portal Sexual Health Instrument by group, American Men's Internet Survey 2018 (N=1375).

Group	AMIS-PPSHI[a]				
	Total				
	Cronbach α	Mean (SD)	Test statistic[b]		*P* value
			F test (df)	*t* test (df)	
All (N=1375)	.8430	28.94 (5.14)	N/A[c]	N/A	N/A
US region			4.37 (3,1371)	N/A	.005
Northeast (eg, New York and Vermont; n=221)	.8208	28.35 (4.88)			
Midwest (eg, Illinois and Ohio; n=314)	.8529	28.29 (5.46)			
South (eg, Florida and Alabama; n=505)	.8461	29.39 (5.13)			
West (eg, California and Oregon; n=335)	.8378	29.24 (4.92)			
Age category			6.11 (3,1371)	N/A	<.001
15-24 (n=318)	.8149	29.52 (4.62)			
25-29 (n=214)	.8095	29.44 (4.57)			
30-39 (n=291)	.8422	29.28 (5.03)			
≥40 (n=552)	.8608	28.22 (5.60)			
STI[d] test in 12 months before study			N/A	1.64 (1373)	.10
Yes (n=1,128)	.8403	29.04 (5.10)			
No (n=247)	.8548	28.45 (5.28)			
History of STI diagnosis			7.89 (2,1372)	N/A	<.001
No history of STI diagnosis (n=1087)	.8385	29.21 (4.98)			
Living with HIV (n=118)	.8617	27.57 (5.89)			
Recent gonorrhea, chlamydia, or syphilis and no HIV (n=170)	.8437	28.13 (5.36)			
Anticipated health care stigma			N/A	0.22 (1373)	.83
No (n=1037)	.8431	28.95 (5.18)			
Yes (n=338)	.8433	28.88 (5.02)			
Serious mental illness			N/A	1.76 (1373)	.08
No (K6[e]<13; n=1113)	.8455	28.82 (5.20)			
Yes (K6≥13; n=262)	.8288	29.44 (4.82)			

[a]American Men's Internet Survey-Patient Portal Sexual Health Instrument.

[b]Test statistic: for analysis of variance (ANOVA) the F-value $_{(\text{degrees of freedom groups, degrees of freedom residuals})}$ test statistic is reported; for *t* tests, the *t* value (degrees of freedom) test statistic is reported.

[c]N/A: not applicable.

[d]STI: sexually transmitted infection.

[e]K6 refers to the Kessler 6-item psychological distress scale.

Table 6. Psychometrics of the Patient Portal Sexual Health Instrument Subscales, by group, American Men's Internet Survey 2018 (N=1375).

Group	AMIS-PPSHI[a] Subscales											
	Sexual health engagement and awareness				Enhancing dyadic communication				Managing sexual health care			
	Mean (SD)	Test statistic[b]		*P* value	Mean (SD)	Test statistics[b]		*P* value	Mean (SD)	Test statistic[b]		*P* value
		F test (df)	*t* test (df)			*F* test (df)	*t* test (df)			*F* test (df)	*t* test (df)	
	7.46 (1.70)	N/A	N/A	N/A	7.57 (1.71)	N/A	N/A	N/A	13.90 (3.22)	N/A	N/A	N/A
US region		2.41 (3,1371)	N/A	.07		0.75 (3,1371)	N/A	.52		5.63 (3,1371)	N/A	<.001
Northeast (eg, New York and Vermont; n=221)	7.36 (1.75)				7.43 (1.68)				13.56 (3.02)			
Midwest (eg, Illinois and Ohio; n=314)	7.30 (1.79)				7.55 (1.77)				13.43 (3.40)			
South (eg, Florida and Alabama; n=505)	7.48 (1.70)				7.63 (1.67)				14.29 (3.25)			
West (eg, California and Oregon; n=335)	7.64 (1.55)				7.60 (1.73)				14.00 (3.06)			
Age category		4.64 (3,1371)	N/A	.003		10.87 (3,1371)	N/A	<.001		1.47 (3,1371)	N/A	.22
15-24 (n=318)	7.51 (1.55)				7.86 (1.38)				14.15 (3.20)			
25-29 (n=214)	7.70 (1.44)				7.78 (1.52)				13.96 (2.99)			
30-39 (n=291)	7.59 (1.65)				7.70 (1.64)				13.99 (3.09)			
≥40 (n=552)	7.26 (1.87)				7.26 (1.93)				13.70 (3.38)			
STI[d] test in 12 months before study		N/A	2.86 (1373)	.004		N/A	1.99 (1373)	.05		N/A	0.09 (1373)	.93
Yes (n=1,128)	7.52 (1.66)				7.62 (1.69)				13.91 (3.23)			
No (n=247)	7.18 (1.83)				7.38 (1.82)				13.89 (3.18)			
History of STI diagnosis		1.02 (2,1372)	N/A	.36		12.25 (2,1372)	N/A	<.001		6.13 (2,1372)	N/A	.002
No history of STI diagnosis (n=1087)	7.49 (1.68)				7.66 (1.64)				14.06 (3.13)			
Living with HIV (n=118)	7.39 (1.72)				6.85 (2.13)				13.33 (3.64)			
Recent gonorrhea, chlamydia, or syphilis and no HIV (n=170)	7.30 (1.78)				7.52 (1.74)				13.31 (3.41)			
Anticipated health care stigma		N/A	0.47 (1373)	.64		N/A	1.02 (1373)	.31		N/A	0.35 (1373)	.73
No (n=1037)	7.47 (1.70)				7.60 (1.74)				13.89 (3.26)			
Yes (n=338)	7.42 (1.67)				7.49 (1.63)				13.96 (3.11)			
Serious mental illness		N/A	0.68 (1373)	.49		N/A	1.79 (1373)	.07		N/A	1.58 (1373)	.11

Group	AMIS-PPSHI[a] Subscales									
	Sexual health engagement and awareness				Enhancing dyadic communication				Managing sexual health care			
	Mean (SD)	Test statistic[b]		*P* value	Mean (SD)	Test statistics[b]		*P* value	Mean (SD)	Test statistic[b]		*P* value
		F test (df)	*t* test (df)			*F* test (df)	*t* test (df)			*F* test (df)	*t* test (df)	
No (K6[e]<13; n=1113)	7.44 (1.72)				7.53 (1.75)				13.84 (3.23)			
Yes (K6≥13; n=262)	7.52 (1.61)				7.74 (1.51)				14.19 (3.17)			

[a]American Men's Internet Survey-Patient Portal Sexual Health Instrument.

[b]Test statistic: for analysis of variance (ANOVA) the F-value $_{(\text{degrees of freedom groups, degrees of freedom residuals})}$ test statistic is reported; for *t* tests, the *t* value (degrees of freedom) test statistic is reported.

[c]N/A: not applicable.

[d]STI: sexually transmitted infection.

[e]K6 refers to the Kessler 6-item psychological distress scale.

Discussion

Principal Findings

The goal of this study is to develop a brief instrument for measuring attitudes focused on using patient portals for STI prevention and care among a nationwide sample of SMM. HIV and other STIs are costly and have a high burden on SMM. Patient portals could be used to address risky sexual behaviors; however, past studies have not looked at service design and consumer adoption models for patient-facing IT solutions. Therefore, we created a brief instrument to enable the measurement of attitudes toward using patient portals for STI prevention and care among SMM with access to a patient portal. The instrument was adapted from earlier scientific work, is short, and may be added to health questionnaires focused on sexual health–related technology use. Furthermore, we demonstrated that AMIS-PPSHI might be adapted to include novel consumer-oriented features as technology evolves.

The resulting instrument consists of 12 items and 3 subscales measuring constructs of (1) sexual health engagement and awareness, (2) enhancing dyadic communication, and (3) managing sexual health care. Constructs cover unique aspects of patient portal use. First, portals should communicate personalized sexual health information to the user. Second, interpersonal use of patient portals may occur outside of health settings, particularly in events that share test histories with sexual partners. Third, patient portals should empower patients to engage with an array of sexual health care management services, such as testing and telehealth.

A Closer Look at PPSHI

PPSHI and its 3 subscales had a strong overall internal consistency. As expected, younger participants are more receptive to technology use. Scores have an indirect relationship with age; SMM aged 15-24 years had the highest scores. Scores stratified by STI diagnosis reveal interesting dynamics, most notably in the enhancing dyadic communication subscale. Scores on the enhancing dyadic communication subscale are the lowest among participants with HIV, an indicator of the highly stigmatized nature of HIV. Interpersonal use of patient portals with sexual partners may likely be lower for participants with HIV or other chronic STIs. Interventions are needed to reduce this stigma and to strengthen self-efficacy for discussing with partners the topics pertinent to sexual health and wellness.

Similar scores by anticipated health care stigma may be an indicator of acceptability of patient portals among participants who may be less likely to receive HIV care services [39,41]. Slightly higher AMIS-PPSHI scores among participants with mental health illness support the acceptability of the patient portal use among patients with mental health disorders [66]. Together with earlier studies, findings support patient portals as a promising avenue to plan interventions around increasing health engagement among marginalized groups, including persons with HIV [28,30,46]. The overall high mean sample score on AMIS-PPSHI may also indicate that patient portal interventions may extend to other areas of health care engagement, such as achieving hepatitis A and B vaccination and screening recommendations for SMM [67-69]. Messages delivered through patient portals have been demonstrated to increase herpes zoster vaccination in adults [70]. Thus, the current global climate of hepatitis A outbreak among SMM reflects missed opportunities to leverage patient portals to deliver hepatitis A and B vaccination screening messages to SMM [71-75]. The application of machine learning algorithms to identify PrEP candidates using EHR data may also be applied to identify candidates for hepatitis A and B vaccination; however, research is needed to develop efficacious algorithms [76].

Strengths and Limitations

The strength of AMIS-PPSHI is that it is based on empirical research, and it is very timely to the growing technology-based STI prevention models. The scale is shortened in item numbers and response options and still holds a strong internal consistency. The factor analysis is based on both an acceptable sample size of participants and the number of observed

variables. Overall, the instrument performed statistically well in psychometric analysis. The age distribution of the sample offers some comparisons and extrapolation across groups of adolescent, young, and adult SMM. However, neither AMIS-PPSHI nor its subscales have been validated as constructs related to patterns of patient portal use for sexual health and wellness. Future validation studies may explore the relationship of constructs with patterns of patient portal use specific to sexual health and wellness. Clinical researchers may test the enhancing dyadic communication subscale as a determinant of an individual's likelihood to use PHRs for sharing STI test histories with main and nonmain partners [16]. The subscale may then be applied in clinical settings with decision analytics to identify patients for interventions that are less likely to disclose STI PHRs with partners. More data are needed on patterns of patient portal use for sexual health services such as viewing electronic STI test results, viewing health information on STIs, and ordering medications to prevent and treat STIs.

African American or Black SMM are notably underrepresented in the study sample. Future studies are needed to apply PPSHI across a broad nationwide sample of Black SMM and youth—the race or ethnic group most overburdened by HIV and other STI incidence and prevalence [34,38]. Additional validation studies are needed among other priority populations for STI prevention in the United States and other countries with burgeoning mHealth environments. A further limitation is that the instrument does not include items focused on the use of patient portals to report sexual health–related behaviors and outcomes. Ecological momentary assessment and patient-reported outcome measures are mechanisms that can feed data into patient portal systems and inform decision support algorithms for the user or health care providers [77-79]. Given the adaptability of PPSHI, items may be added to patient-reported outcomes to assess perceptions about reporting personal data to the patient portal for sexual health.

Conclusions

In summary, we suggest that PPSHI and its components could predict behavioral intentions and patterns for patient portal use for health behaviors related to STI prevention and care among SMM. PPSHI is feasibly adaptable to questionnaires and may have useful applications in electronic patient intake surveys. Short surveys on patient intake forms assessing risk behaviors have been used to inform clinical decision support algorithms, prompting providers to encourage STI screening for patients [80]. Assessing PPSHI constructs in patients may similarly be useful in informing provider messaging within decision support algorithms, for example, encouraging patients to share STI PHRs with partners.

Acknowledgments

The authors thank Dr Stefan Baral at Johns Hopkins University (Baltimore, MD, USA) and Dr Travis Sanchez at Emory University (Atlanta, GA, USA) for generously creating space for this research in the 2018 American Men's Internet Survey. This work was supported by grants from the MAC AIDS Fund and by the National Institutes of Health (P30AI050409), the Emory Center for AIDS Research, and the Johns Hopkins Bloomberg School of Public Health Center for Public Health and Human Rights (5R01MH110358-03). KJ's efforts were supported, in part, by T32DA007292 and R25MH08362.

Authors' Contributions

All listed authors contributed meaningfully to manuscript writing and revision in preparation for peer review submission. KJ analyzed the data and drafted the paper as the first author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Electronic Sexual Health Information Notification and Education (eSHINE) Study Patient Portal Instrument, 2014-2016, N=35.
[DOCX File , 17 KB - jmir_v23i2e18750_app1.docx]

References

1. Horvath KJ, Alemu D, Danh T, Baker JV, Carrico AW. Creating Effective Mobile Phone Apps to Optimize Antiretroviral Therapy Adherence: Perspectives From Stimulant-Using HIV-Positive Men Who Have Sex With Men. JMIR mHealth uHealth 2016 Apr 15;4(2):e48. [doi: 10.2196/mhealth.5287]
2. Horvath KJ, Lammert S, MacLehose RF, Danh T, Baker JV, Carrico AW. A Pilot Study of a Mobile App to Support HIV Antiretroviral Therapy Adherence Among Men Who Have Sex with Men Who Use Stimulants. AIDS Behav 2019 Jul 15;23(11):3184-3198. [doi: 10.1007/s10461-019-02597-3]
3. Horvath KJ, Amico KR, Erickson D, Ecklund AM, Martinka A, DeWitt J, et al. Thrive With Me: Protocol for a Randomized Controlled Trial to Test a Peer Support Intervention to Improve Antiretroviral Therapy Adherence Among Men Who Have Sex With Men. JMIR Res Protoc 2018 May 31;7(5):e10182. [doi: 10.2196/10182]

4. Bauermeister JA, Golinkoff JM, Horvath KJ, Hightow-Weidman LB, Sullivan PS, Stephenson R. A Multilevel Tailored Web App–Based Intervention for Linking Young Men Who Have Sex With Men to Quality Care (Get Connected): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2018 Aug 02;7(8):e10444. [doi: 10.2196/10444]
5. Office of the National Coordinator for Health Information Technology (ONC). Individuals' Perceptions of the Privacy and Security of Medical Records and Health Information Exchange. Health IT Quick-Stat #58; Accessed 02/16/2020, 2020. 2019. URL: https://dashboard.healthit.gov/quickstats/pages/consumers-privacy-security-medical-record-information-exchange.php [accessed 2021-01-09]
6. Anderberg P, Eivazzadeh S, Berglund JS. A Novel Instrument for Measuring Older People's Attitudes Toward Technology (TechPH): Development and Validation. J Med Internet Res 2019 May 23;21(5):e13951. [doi: 10.2196/13951]
7. Kim J, Park H. Development of a Health Information Technology Acceptance Model Using Consumers' Health Behavior Intention. J Med Internet Res 2012 Oct 01;14(5):e133. [doi: 10.2196/jmir.2143]
8. ONC. What is a personal health record? Healthit.gov. 2016. URL: https://www.healthit.gov/faq/what-personal-health-record-0 [accessed 2021-01-09]
9. ONC. What is a patient portal? Healthit.gov. 2017. URL: https://www.healthit.gov/faq/what-patient-portal [accessed 2021-01-09]
10. ONC. What is Health IT? Healthit.gov. 2019. URL: https://www.healthit.gov/faq/what-health-it [accessed 2021-01-09]
11. ONC. About The ONC Health IT Certification Program. Healthit.gov. 2020. URL: https://www.healthit.gov/topic/certification-ehrs/about-onc-health-it-certification-program [accessed 2021-01-09]
12. ONC. 2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications. Federal Register 2015;80(200):16804-16921. [doi: 10.1007/springerreference_82314] [Medline: 26477063]
13. Jackman KP, Hightow-Weidman L, Poteat T, Wirtz AL, Kane JC, Baral SD. Evaluating psychometric determinants of willingness to adopt sexual health patient portal services among black college students: A mixed-methods approach. Journal of American College Health 2019 Nov 11:1-8. [doi: 10.1080/07448481.2019.1660352] [Medline: 31710578]
14. Hoque MR, Bao Y, Sorwar G. Investigating factors influencing the adoption of e-Health in developing countries: A patient's perspective. Informatics for Health and Social Care 2016 Feb 11;42(1):1-17. [doi: 10.3109/17538157.2015.1075541] [Medline: 26865037]
15. Jackman K, Latkin CA, Maksut JL, Trent ME, Sanchez TH, Baral SD. Patient Portals as Highly Acceptable Tools to Support HIV Preventative Behaviors Among Adolescent and Young Sexual Minority Men. Journal of Adolescent Health 2020 Aug;67(2):278-281. [doi: 10.1016/j.jadohealth.2020.03.029] [Medline: 32387095]
16. Jackman KM, Dangerfield DT, Yang C, Trent M, Kharrazi H, Johnson RM, et al. Behavioral intentions to use patient portals to disclose HIV and other sexually transmitted infection testing histories with sexual partners among US sexual minority men. AIDS Behav 2020 Nov 13:-. [doi: 10.1007/s10461-020-03092-w] [Medline: 33185776]
17. ONC. Certified Health IT Developers and Editions Reported by Hospitals Participating in the Medicare EHR Incentive Program, Health IT Quick-Stat #29. Healthit.gov. 2017. URL: https://dashboard.healthit.gov/quickstats/pages/FIG-Vendors-of-EHRs-to-Participating-Hospitals.php [accessed 2021-01-09]
18. ONC. §170.315(g)(8) Application access — data category request. Healthit.gov. 2020. URL: https://www.healthit.gov/test-method/application-access-data-category-request [accessed 2021-01-09]
19. ONC. Number of 2015 Edition Unique Products certified to specific Certification Criteria. Healthit.gov. URL: https://chpl.healthit.gov/#/charts [accessed 2020-01-09]
20. ONC. Advancing Care Information Reporting: Successfully Report on Your MACRA Advancing Care Information Measures Using Certified Health IT. Healthit.gov. 2019. URL: https://www.healthit.gov/topic/federal-incentive-programs/MACRA/MIPS/advancing-care-information-reporting [accessed 2021-01-09]
21. Ford EW, Hesse BW, Huerta TR. Personal Health Record Use in the United States: Forecasting Future Adoption Levels. J Med Internet Res 2016 Mar 30;18(3):e73. [doi: 10.2196/jmir.4973]
22. ONC. U.S. Hospital Adoption of Patient Engagement Functionalities. Healthit.gov. 2016. URL: https://dashboard.healthit.gov/quickstats/pages/FIG-Hospital-Adoption-of-Patient-Engagement-Functionalities.php [accessed 2021-01-09]
23. ONC. Office-based Physician Electronic Patient Engagement Capabilities. Healthit.gov. 2016. URL: dashboard.healthit.gov/quickstats/pages/physicians-view-download-transmit-secure-messaging-patient-engagement.php [accessed 2021-01-09]
24. Centers for Medicare and Medicaid Services. CLIA program and HIPAA privacy rule; patients' access to test reports. Federal register 2014;79:7289-7316. [doi: 10.1201/b17548-8]
25. Zhao J, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, Facilitators, and Solutions to Optimal Patient Portal and Personal Health Record Use: A Systematic Review of the Literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
26. Emani S, Yamin CK, Peters E, Karson AS, Lipsitz SR, Wald JS, et al. Patient Perceptions of a Personal Health Record: A Test of the Diffusion of Innovation Model. J Med Internet Res 2012 Nov 05;14(6):e150. [doi: 10.2196/jmir.2278]
27. King G, Maxwell J, Karmali A, Hagens S, Pinto M, Williams L, et al. Connecting Families to Their Health Record and Care Team: The Use, Utility, and Impact of a Client/Family Health Portal at a Children's Rehabilitation Hospital. J Med Internet Res 2017 Apr 06;19(4):e97. [doi: 10.2196/jmir.6811]

28. Jackman K, Latkin CA, Maksut JL, Trent ME, Sanchez TH, Baral SD. Patient Portals as Highly Acceptable Tools to Support HIV Preventative Behaviors Among Adolescent and Young Sexual Minority Men. Journal of Adolescent Health 2020 Aug;67(2):278-281. [doi: 10.1016/j.jadohealth.2020.03.029]
29. Jackman KP, Hightow-Weidman L, Poteat T, Wirtz AL, Kane JC, Baral SD. Evaluating psychometric determinants of willingness to adopt sexual health patient portal services among black college students: A mixed-methods approach. Journal of American College Health 2019 Nov 11:1-8. [doi: 10.1080/07448481.2019.1660352]
30. Jackman K, Baral SD, Hightow-Weidman L, Poteat T. Uncovering a Role for Electronic Personal Health Records in Reducing Disparities in Sexually Transmitted Infection Rates Among Students at a Predominantly African American University: Mixed-Methods Study. JMIR Med Inform 2018 Jul 12;6(3):e41. [doi: 10.2196/medinform.9174]
31. CDC. Reported STDs in the United States, 2016 High Burden of STDs Threaten Millions of Americans. CDC. 2017. URL: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwik8O7Vh4_uAhXrwVkKHT_NCH0QFjABegQIARAC&url=https%3A%2F%2Fstacks.cdc.gov%2Fview%2Fcdc%2F48599%2Fcdc_48599_DS1.pdf%3F&usg=AOvVaw3iT_Ry7mbDSGUYvox3OCvs [accessed 2021-01-09]
32. CDC. New CDC Report: STDs Continue to Rise in the U.S. NCHHSTP Newsroom 2019; Accessed 07/24/2020. NCHHSTP Newsroom. 2019. URL: https://www.cdc.gov/nchhstp/newsroom/2019/2018-STD-surveillance-report-press-release.html [accessed 2021-01-09]
33. Młynarczyk-Bonikowska B, Majewska A, Malejczyk M, Młynarczyk G, Majewski S. Multiresistant Neisseria gonorrhoeae: a new threat in second decade of the XXI century. Med Microbiol Immunol 2019 Dec 04;209(2):95-108. [doi: 10.1007/s00430-019-00651-4]
34. CDC. Sexually Transmitted Disease Surveillance 2018. CDC. Atlanta: U.S. Department of Health and Human Services; 2019. URL: https://www.cdc.gov/std/stats18/STDSurveillance2018-full-report.pdf [accessed 2021-01-09]
35. CDC. Preexposure prophylaxis for the prevention of HIV infection in the United States - 2017 Update. CDC 2018:1-77.
36. National Institute of Allergy and Infectious Disease. HIV Undetectable=Untransmittable (U=U), or Treatment as Prevention. naid.nih.gov. 2019. URL: https://www.niaid.nih.gov/diseases-conditions/treatment-prevention [accessed 2021-01-09]
37. CDC. Gay, Bisexual,Other MSM. CDC. 2015. URL: https://www.cdc.gov/std/life-stages-populations/msm.htm [accessed 2021-01-09]
38. CDC. HIV Surveillance Report, 2018 (Updated). CDC. 2020. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2021-01-09]
39. Golub SA, Gamarel KE. The Impact of Anticipated HIV Stigma on Delays in HIV Testing Behaviors: Findings from a Community-Based Sample of Men Who Have Sex with Men and Transgender Women in New York City. AIDS Patient Care and STDs 2013 Nov;27(11):621-627. [doi: 10.1089/apc.2013.0245]
40. Fisher CB, Fried AL, Macapagal K, Mustanski B. Patient–Provider Communication Barriers and Facilitators to HIV and STI Preventive Services for Adolescent MSM. AIDS Behav 2018 Mar 15;22(10):3417-3428. [doi: 10.1007/s10461-018-2081-x]
41. Sullivan KA, Berger MB, Quinlivan EB, Parnell HE, Sampson LA, Clymore JM, et al. Perspectives from the Field. Journal of the International Association of Providers of AIDS Care (JIAPAC) 2016 Jul 09;15(6):477-485. [doi: 10.1177/2325957415617830]
42. Vreeman RC, McCoy BM, Lee S. Mental health challenges among adolescents living with HIV. Journal of the International AIDS Society 2017 May 16;20:21497. [doi: 10.7448/ias.20.4.21497]
43. Gokhale RH, Weiser J, Sullivan PS, Luo Q, Shu F, Bradley H. Depression Prevalence, Antidepressant Treatment Status, and Association with Sustained HIV Viral Suppression Among Adults Living with HIV in Care in the United States, 2009–2014. AIDS Behav 2019 Jul 31;23(12):3452-3459. [doi: 10.1007/s10461-019-02613-6]
44. Beyrer C, Sullivan PS, Sanchez J, Dowdy D, Altman D, Trapence G, et al. A call to action for comprehensive HIV services for men who have sex with men. The Lancet 2012 Jul;380(9839):424-438. [doi: 10.1016/s0140-6736(12)61022-8]
45. Sullivan PS, Driggers R, Stekler JD, Siegler A, Goldenberg T, McDougal SJ, et al. Usability and Acceptability of a Mobile Comprehensive HIV Prevention App for Men Who Have Sex With Men: A Pilot Study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e26. [doi: 10.2196/mhealth.7199]
46. Luque AE, van Keken A, Winters P, Keefer MC, Sanders M, Fiscella K. Barriers and Facilitators of Online Patient Portals to Personal Health Records Among Persons Living With HIV: Formative Research. JMIR Res Protoc 2013 Jan 22;2(1):e8. [doi: 10.2196/resprot.2302]
47. Hoque MR, Bao Y, Sorwar G. Investigating factors influencing the adoption of e-Health in developing countries: A patient's perspective. Informatics for Health and Social Care 2016 Feb 11;42(1):1-17. [doi: 10.3109/17538157.2015.1075541]
48. Tavares J, Oliveira T. New Integrated Model Approach to Understand the Factors That Drive Electronic Health Record Portal Adoption: Cross-Sectional National Survey. J Med Internet Res 2018 Nov 19;20(11):e11032. [doi: 10.2196/11032]
49. Sanchez TH, Sineath RC, Kahle EM, Tregear SJ, Sullivan PS. The Annual American Men's Internet Survey of Behaviors of Men Who Have Sex With Men in the United States: Protocol and Key Indicators Report 2013. JMIR Public Health Surveill 2015 Apr 17;1(1):e3. [doi: 10.2196/publichealth.4314]
50. Jackman KM, Murray S, Hightow-Weidman L, Trent ME, Wirtz AL, Baral SD, et al. Digital technology to address HIV and other sexually transmitted infection disparities: intentions to disclose online personal health records to sex partners

among students at a historically Black college. PLoS One 2020;15(8):e0237648 [FREE Full text] [doi: 10.1371/journal.pone.0237648] [Medline: 32822360]
51. Wootton AR, Legnitto DA, Gruber VA, Dawson-Rose C, Neilands TB, Johnson MO, et al. Telehealth and texting intervention to improve HIV care engagement, mental health and substance use outcomes in youth living with HIV: a pilot feasibility and acceptability study protocol. BMJ Open 2019 Jul 16;9(7):e028522. [doi: 10.1136/bmjopen-2018-028522]
52. Sullivan S, Sullivan P, Stephenson R. Acceptability and Feasibility of a Telehealth Intervention for Sexually Transmitted Infection Testing Among Male Couples: Protocol for a Pilot Study. JMIR Res Protoc 2019 Oct 1;8(10):e14481. [doi: 10.2196/14481]
53. Muessig K, Knudston K, Soni K. 'I didn't tell you sooner because I didn't know how to handle it myself.' developing a virtual reality program to support HIV-status disclosure decisions. Digit Cult Educ 2018;10:22-48. [Medline: 30123342]
54. Grossman L, Masterson Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019:855-870. [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
55. Muessig K, LeGrand S, Horvath K, Bauermeister J, Hightow-Weidman L. Recent mobile health interventions to support medication adherence among HIV-positive MSM. Current Opinion in HIV and AIDS 2017;12(5):432-441. [doi: 10.1097/coh.0000000000000401] [Medline: 28639990]
56. Liu A, Coleman K, Bojan K, Serrano PA, Oyedele T, Garcia A, et al. Developing a Mobile App (LYNX) to Support Linkage to HIV/Sexually Transmitted Infection Testing and Pre-Exposure Prophylaxis for Young Men Who Have Sex With Men: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Jan 25;8(1):e10659. [doi: 10.2196/10659] [Medline: 30681964]
57. LeGrand S, Muessig K, Horvath K, Rosengren A, Hightow-Weidman L. Using technology to support HIV self-testing among MSM. Current Opinion in HIV and AIDS 2017;12(5):425-431. [doi: 10.1097/coh.0000000000000400] [Medline: 28617712]
58. Lu AS, Kharrazi H. A State-of-the-Art Systematic Content Analysis of Games for Health. Games for Health Journal 2018 Feb;7(1):1-15. [doi: 10.1089/g4h.2017.0095] [Medline: 29293368]
59. Franklin S, Gibson D, Robertson P, Pohlmann J, Fralish J. Parallel Analysis: a Method for Determining Significant Principal Components Recommended Citation. Journal of Vegetation Science 1995;6(1):99-106. [doi: 10.2307/3236261]
60. Kaiser HF. An index of factorial simplicity. Psychometrika 1974 Mar;39(1):31-36. [doi: 10.1007/bf02291575]
61. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
62. Hair J, Black W, Babin B, Anderson R. Multivariate Data Analysis (seventh edition). Upper Saddle River, NJ: Prentice Hall; 2019.
63. Kessler RC, Green JG, Gruber MJ, Sampson NA, Bromet E, Cuitan M, et al. Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. Int. J. Methods Psychiatr. Res 2011 Feb 24;20(1):62-62. [doi: 10.1002/mpr.333]
64. Brown JD. Choosing the Right Type of Rotation in PCA and EFA. Shiken: JALT Testing & Evaluation SIG Newsletter 2009;13(3):20-25.
65. DiStefano C, Zhu M, Mindrila D. Understanding and Using Factor Scores: Considerations for the Applied Researcher. Practical Assessment, Research & Evaluation 2009;14(20):1-11.
66. Saberi P, Dawson Rose C, Wootton AR, Ming K, Legnitto D, Jeske M, et al. Use of technology for delivery of mental health and substance use services to youth living with HIV: a mixed-methods perspective. AIDS Care 2019 May 28;32(8):931-939. [doi: 10.1080/09540121.2019.1622637]
67. Advisory Committee on Immunization Practices (ACIP), Fiore AE, Wasley A, Bell BP. Prevention of hepatitis A through active or passive immunization: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR Recomm Rep 2006 May 19;55(RR-7):1-23 [FREE Full text] [Medline: 16708058]
68. Schillie S, Vellozzi C, Reingold A, Harris A, Haber P, Ward JW, et al. Prevention of Hepatitis B Virus Infection in the United States: Recommendations of the Advisory Committee on Immunization Practices. MMWR Recomm. Rep 2018 Jan 12;67(1):1-31. [doi: 10.15585/mmwr.rr6701a1]
69. Weinbaum CM, Mast EE, Ward JW. Recommendations for identification and public health management of persons with chronic hepatitis B virus infection. Hepatology 2009 Apr 27;49(S5):S35-S44. [doi: 10.1002/hep.22882]
70. Otsuka SH, Tayal NH, Porter K, Embi PJ, Beatty SJ. Improving Herpes Zoster Vaccination Rates Through Use of a Clinical Pharmacist and a Personal Health Record. Am J Med 2013 Sep;126(9):832.e1-832.e6. [doi: 10.1016/j.amjmed.2013.02.018]
71. Cohall A, Zucker J, Krieger R, Scott C, Guido C, Hakala S, et al. Missed Opportunities for Hepatitis A Vaccination Among MSM Initiating PrEP. J Community Health 2019 Nov 13;45(3):506-509. [doi: 10.1007/s10900-019-00768-w]
72. Foster MA, Hofmeister MG, Kupronis BA, Lin Y, Xia G, Yin S, et al. Increase in Hepatitis A Virus Infections — United States, 2013–2018. MMWR Morb Mortal Wkly Rep 2019 May 10;68(18):413-415. [doi: 10.15585/mmwr.mm6818a2]
73. Tram J, Le Baccon‐Sollier P, Bolloré K, Ducos J, Mondain A, Pastor P, et al. RNA testing for the diagnosis of acute hepatitis A during the 2017 outbreak in France. J Viral Hepat 2020 May;27(5):540-543. [doi: 10.1111/jvh.13255]

74. Raczyńska A, Wickramasuriya NN, Kalinowska-Nowak A, Garlicki A, Bociąga-Jasik M. Acute Hepatitis A Outbreak Among Men Who Have Sex With Men in Krakow, Poland; February 2017–February 2018. Am J Mens Health 2019 Dec 26;13(6):155798831989514. [doi: 10.1177/1557988319895141]
75. Sachdeva H, Benusic M, Ota S, Stuart R, Maclachlan J, Dubey V, et al. Community outbreak of hepatitis A disproportionately affecting men who have sex with men in Toronto, Canada, January 2017–November 2018. Can Commun Dis Rep 2019 Oct 3;45(10):262-268. [doi: 10.14745/ccdr.v45i10a03]
76. Marcus JL, Hurley LB, Krakower DS, Alexeeff S, Silverberg MJ, Volk JE. Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: a modelling study. The Lancet HIV 2019 Oct;6(10):e688-e695. [doi: 10.1016/s2352-3018(19)30137-7]
77. Yang C, Linas B, Kirk G, Bollinger R, Chang L, Chander G, et al. Feasibility and Acceptability of Smartphone-Based Ecological Momentary Assessment of Alcohol Use Among African American Men Who Have Sex With Men in Baltimore. JMIR mHealth uHealth 2015 Jun 17;3(2):e67. [doi: 10.2196/mhealth.4344]
78. Cronin R, Conway D, Condon D, Jerome R, Byrne D, Harris P. Patient and healthcare provider views on a patient-reported outcomes portal. J Am Med Inform Assoc 2018 Nov 01;25(11):1470-1480. [doi: 10.1093/jamia/ocy111] [Medline: 30239733]
79. Wu AW, Kharrazi H, Boulware LE, Snyder CF. Measure once, cut twice—adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. Journal of Clinical Epidemiology 2013 Aug;66(8):S12-S20. [doi: 10.1016/j.jclinepi.2013.04.005]
80. Goyal MK, Fein JA, Badolato GM, Shea JA, Trent ME, Teach SJ, et al. A Computerized Sexual Health Survey Improves Testing for Sexually Transmitted Infection in a Pediatric Emergency Department. J Pediatr 2017 Apr;183:147-152.e1. [doi: 10.1016/j.jpeds.2016.12.045]

Abbreviations

AMIS: American Men's Internet Survey
CFA: confirmatory factor analysis
CFI: comparative fit index
EFA: exploratory factor analysis
EHR: electronic health record
eSHINE: Electronic Sexual Health Information Notification and Education
Health IT: Health information technology
HINTS: Health Information National Trends Survey
KMO: Kaiser-Meyer-Olkin
K6: Kessler 6-item psychological distress scale
mHealth: mobile health
ONC: Office of the National Coordinator for Health Information Technology
PCA: principal component analysis
PHR: personal health record
PPI: patient portal instrument
PPSHI: patient portal sexual health instrument
PrEP: pre-exposure prophylaxis
RMSEA: root mean square error of approximation
SMM: sexual minority men
SRMR: standardized root mean square residual
STD: sexually transmitted disease
STI: sexually transmitted infection
TLI: Tucker-Lewis index

Edited by G Eysenbach; submitted 16.03.20; peer-reviewed by S Naar, G Strudwick; comments to author 17.06.20; revised version received 24.07.20; accepted 26.10.20; published 10.02.21.

Please cite as:
Jackman KMP, Kane J, Kharrazi H, Johnson RM, Latkin C
Using the Patient Portal Sexual Health Instrument in Surveys and Patient Questionnaires Among Sexual Minority Men in the United States: Cross-sectional Psychometric Validation Study
J Med Internet Res 2021;23(2):e18750
URL: http://www.jmir.org/2021/2/e18750/
doi:10.2196/18750
PMID:33565987

Original Paper

Online Medical Record Nonuse Among Patients: Data Analysis Study of the 2019 Health Information National Trends Survey

Safa Elkefi[1], MSc; Zhongyuan Yu[1], PhD; Onur Asan[1], PhD

School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, United States

Corresponding Author:
Onur Asan, PhD
School of Systems and Enterprises
Stevens Institute of Technology
1 Castle Point Terrace
Hoboken, NJ, 07030
United States
Phone: 1 4145264330
Email: oasan@stevens.edu

Abstract

Background: Online medical records are being used to organize processes in clinical and outpatient settings and to forge doctor-patient communication techniques that build mutual understanding and trust.

Objective: We aimed to understand the reasons why patients tend to avoid using online medical records and to compare the perceptions that patients have of online medical records based on demographics and cancer diagnosis.

Methods: We used data from the Health Information National Trends Survey Cycle 3, a nationally representative survey, and assessed outcomes using descriptive statistics and chi-square tests. The patients (N=4328) included in the analysis had experienced an outpatient visit within the previous 12 months and had answered the online behavior question regarding their use of online medical records.

Results: Patients who were nonusers of online medical records consisted of 58.36% of the sample (2526/4328). The highest nonuser rates were for patients who were Hispanic (460/683, 67.35%), patients who were non-Hispanic Black (434/653, 66.46%), and patients who were older than 65 years (968/1520, 63.6%). Patients older than 65 years were less likely to use online medical records (odds ratio [OR] 1.51, 95% CI 1.24-1.84, *P*<.001). Patients who were White were more likely to use online medical records than patients who were Black (OR 1.71, 95% CI 1.43-2.05, *P*<.001) or Hispanic (OR 1.65, 95% CI 1.37-1.98, *P*<.001). Patients who were diagnosed with cancer were more likely to use online medical records compared to patients with no cancer (OR 1.31, 95% CI 1.11-1.55, 95% CI 1.11-1.55, *P*=.001). Among nonusers, older patients (≥65 years old) preferred speaking directly to their health care providers (OR 1.76, 95% CI 1.35-2.31, *P*<.001), were more concerned about privacy issues caused by online medical records (OR 1.79, 95% CI 1.22-2.66, *P*<.001), and felt uncomfortable using the online medical record systems (OR 10.55, 95% CI 6.06-19.89, *P*<.001) compared to those aged 18-34 years. Patients who were Black or Hispanic were more concerned about privacy issues (OR 1.42, 1.09-1.84, *P*=.007).

Conclusions: Studies should consider social factors such as gender, race/ethnicity, and age when monitoring trends in eHealth use to ensure that eHealth use does not induce greater health status and health care disparities between people with different backgrounds and demographic characteristics.

(J Med Internet Res 2021;23(2):e24767) doi:10.2196/24767

KEYWORDS

online medical records; cancer; patient portals; communication; medical records; health information; eHealth

Introduction

Communication is a key element in providing high-quality health care services [1]. If communication is effective, it can lead to significant health care provision and positive outcomes, including decreased anxiety, guilt, pain, and disease symptoms. Moreover, effective communication can increase patient satisfaction, acceptance, adherence, and cooperation with the medical team, and it can improve the physiological and functional status of the patient [1]. Conversely, poor communication between doctors and patients can lead to poor

quality and continuity of care [2]. Patients with communication-related disabilities or problems were 3 times more likely to experience medical or clinical complications compared to other patients [3]. Thus, ensuring good communication by recording, handling, and sharing health information with patients remains a necessity and an integral component of the health care process that may improve outcomes and limit or prevent costly duplication of tests and treatments [4,5].

Information exchange is not only limited to the visit. It is moving beyond that, as patients have the opportunity to access their information regardless of location or time via technologies [4]. Technological advancement has provided opportunities not only to improve doctor-patient information exchange but also to inform and empower patients' role in decision making. Information and communication technology is often used for communication and information exchange purposes and may be a valuable tool for handling these challenges [6,7]. Other potential solutions include means for patients to access their online medical records, including personalized health records [8-11] and patient portals [12]. These centralized health information systems could also have the greatest benefit for information transmission. Although patient portals such as MyPreventiveCare [13], Ask an Expert [14], and others are considered important tools for the development of patient-centered care, their current use is not optimal, and portals are still less patient-centered than they could be [15-17]. Known barriers to the use of portals for patients and providers include security and privacy concerns, the potential negative impact on provider workflow, and limited user friendliness [12,18].

The ability of patients or individuals to access their online medical records serves as one of the backbones to improve patient engagement in and the outcomes of our health care system. Historically, patients have had low access to online medical records. For instance, only 3 out of 10 patients were offered access to medical records in 2013, and almost half of those offered access viewed their online records at least once [19]. However, the use of online medical records has experienced significant growth in recent years, and technological improvements have been made to improve usability and implementation [20]. On the other hand, despite greater availability, there are many patients who still avoid using online medical records, though this may be changing with the current COVID-19 pandemic paradigm shift. The purpose of this study was to explore factors leading to the use or nonuse of online medical records across different groups of patients.

Methods

Data Source

Data for this study were derived from the National Cancer Institute's 2019 Health Information National Trends Survey (HINTS). HINTS is a nationally representative survey (of the US noninstitutionalized adult population) that collects data on the American public's need for, access to, and use of health-related information [21]. HINTS is publicly available on the web [22]. Data used in this study were from HINTS 5 Cycle 3, collected between January and May of 2019. Patients who gave information about their online medical records use were included (survey response rate: 4328/14332, 30.20%). Further details on survey design and sampling strategies are published elsewhere [23].

Study Variables

The following questions correspond to the measures used in the analysis of the study. The first question was, "How many times did you access your online medical record in the last 12 months?" We used this question to identify users and nonusers of online medical records. The respondents who reported accessing their medical record at least once were coded as *users* and respondents who reported accessing their records zero times were coded as *nonusers*. The primary population of interest of the study was nonusers. We specifically analyzed the responses of nonusers to the following questions (Table 1) regarding the reasons why they do not use online medical records.

Table 1. Questions for nonusers and corresponding factors.

Factor name	Question[a]
SpeakDirectly	You prefer to speak to your health care provider directly?
NoNeed	You did not have a need to use your online medical record?
ConcernedPrivacy	You were concerned about the privacy or security of your medical records' website?
NoRecord	You do not have an online medical record?
LogInProb	You found it difficult to login (for example, you had trouble remembering your password)?
Uncomfortable	You are not comfortable or experienced with computers?
MultipleRec	You have more than one online medical record?

[a]All questions had binary (yes or no) responses.

Data Analysis

Descriptive statistics for the HINTS 5 Cycle 3 populations were generated for demographic variables (gender, age, race/ethnicity) and the cancer diagnosis variable. We investigated the relationship between the use of online medical records and the demographic or diagnosis variables using an unadjusted model—Fisher exact test. We focused on patients who had not used online medical records to better understand the reasons behind their avoidance of this particular technology. We report odds ratios (OR) and corresponding 95% confidence intervals; statistical significance was determined based on $P<.05$. All

analyses were conducted using R statistical software (version 3.6.3; packages: Ime4 and Stats).

Results

Sample Characteristics

HINTS 5 Cycle 3 had a total of 4328 respondents, and 57.74% of the respondents (2499/4328) were female. Most respondents were older than 50 years (2926/4328, 67.61%) and non-Hispanic White (2992/4328, 69.13%), and 16% of respondents (693/4328) had been diagnosed with cancer (Table 2).

Table 2. Demographic characteristics of the patients.

Characteristic	Respondents (N=4328), n (%)
Gender	
Male	1829 (42.26)
Female	2499 (57.74)
Age (years)	
18-34	580 (13.40)
35-49	822 (18.99)
50-64	1406 (32.49)
≥65	1520 (35.12)
Race	
Hispanic	683 (15.78)
Non-Hispanic White	2992 (69.13)
Non-Hispanic Black	653 (15.09)
Cancer diagnosis	
Have cancer	693 (16.01)
No cancer	3635 (83.99)

Use of Online Medical Records

Among the 4328 respondents, 2526 (58.36%) were nonusers. The ratio of nonusers across demographics ranged between 53% to 67%, with patients who were Hispanic (460/683, 67.35%) and patients who were non-Hispanic Black being highest (434/653, 66.46%). Patients who were non-Hispanic White (1360/2992, 45.45%) and patients between 18 and 34 years of age had the highest online medical record use (269/580, 46.38%). Table 3 presents percentages of users and nonusers across demographics and nonuser comparisons.

Table 3. Use of online medical records.

Characteristics	All, n (%)	Nonuser, n (%)	User, n (%)	Odds ratio (95% CI)	*P* value
All	4328	2526 (58.36)	1802 (41.64)	N/A[a]	N/A
Gender					
Male	1829 (42.26)	1130 (61.78)	699 (38.22)	1	N/A
Female	2499 (57.74)	1396 (55.86)	1103 (44.14)	0.78 (0.69-0.88)	<.001
Age (years)					
18-34	580 (13.40)	311 (53.62)	269 (46.38)	1	N/A
35-49	822 (18.99)	450 (54.74)	372 (45.26)	1.04 (0.84-1.30)	
50-64	1406 (32.49)	797 (56.69)	609 (43.31)	1.13 (0.92-1.38)	
≥65	1520 (35.12)	968 (63.68)	552 (36.32)	1.51 (1.24-1.84)	<.001
Race					
Non-Hispanic White	2992 (69.13)	1632 (54.55)	1360 (45.45)	1	N/A
Hispanic	683 (15.78)	460 (67.35)	223 (32.65)	1.71 (1.43-2.05)	<.001
Non-Hispanic Black	653 (15.09)	434 (66.46)	219 (33.54)	1.65 (1.37-1.98)	<.001
Cancer diagnosis					
Have cancer	693 (16.01)	365 (52.67)	328 (47.33)	1	N/A
No cancer	3635 (83.99)	2161 (59.95)	1474 (40.05)	1.31 (1.11-1.55)	<.001

[a]N/A: not applicable.

Patients who avoided using online medical records were more likely to be male (OR 0.78, 95% CI 0.69-0.88, *P*<.001). The oldest patients (aged >65 years) were less likely to use online medical records (OR 1.51, 95% CI 1.24-1.84, *P*<.001). Patients who were non-Hispanic White were more likely to use online medical records than patients who were Hispanic (OR 1.71, 95% CI 1.43-2.05, *P*<.001) or non-Hispanic Black (OR 1.65, 95% CI 1.37-1.98, *P*<.001). Finally, patients diagnosed with cancer were more likely to use online medical records than patients without cancer (OR 1.31, 95% CI 1.11-1.55, *P*=.001).

Reasons for Nonuse of Online Medical Records

In the second phase of the analysis, we only focused on nonusers (n=2526) and explored factors regarding their preference for not using online medical records. We compared the different demographics for each factor. The survey had 7 listed factors asked to each respondent. Each participant responded to each question with “yes” or “no.” We summarized the percentage for each demographic group among nonusers. The “desire to speaking directly to the health care provider” was the primary factor (1575/2526, 62.35%) influencing the nonuse of online medical records. Across all demographic characteristics, more than half of respondents answered “yes” for this question. For instance, 63.68% (889/1396) of female nonusers reported “desire to speaking directly” as one of the primary reasons. Almost half of the participants also expressed “no need” as a reason to avoid online medical records. The 18-34 years age group of nonusers had the highest rate of “no need” factor to explain their avoidance of online medical record use (194/311, 62.38%). Privacy concerns were not a primary reason to avoid online medical record use across all groups (range 12% to 23%). Table 4 illustrates all descriptive statistics for the overall nonuser population and each demographic group.

Table 4. Frequencies of reasons that explain nonuse of online medical records.

Characteristic	Factor, n (%)						
	SpeakDirectly	NoNeed	ConcernedPrivacy	NoRecord	LogInProb	Uncomfortable	MultipleRec
All	1575 (62.35)	1258 (49.80)	499 (19.75)	571 (22.60)	435 (17.22)	547 (21.65)	228 (9.03)
Gender							
Male	686 (60.71)	594 (52.57)	214 (18.94)	288 (25.49)	170 (15.04)	253 (22.39)	109 (9.65)
Female	889 (63.68)	664 (47.56)	285 (20.42)	283 (20.27)	265 (18.98)	294 (21.06)	119 (8.52)
Age (years)							
18-34	166 (53.38)	194 (62.38)	39 (12.54)	93 (29.90)	45 (14.47)	14 (4.50)	26 (8.36)
35-49	231 (51.33)	252 (56.00)	83 (18.44)	114 (25.33)	71 (15.78)	48 (10.67)	35 (7.78)
50-64	530 (66.50)	386(48.43)	179 (22.46)	173 (21.71)	130 (16.31)	163 (20.45)	77 (9.66)
≥65	648 (66.94)	426 (44.01)	198 (20.45)	191 (19.73)	189 (19.52)	322 (33.26)	90 (9.30)
Race							
Non-Hispanic White	1004 (61.52)	922 (56.50)	289 (17.71)	375 (22.98)	294 (18.01)	334 (20.47)	164 (10.05)
Hispanic	294 (63.91)	172 (37.39)	108 (23.48)	100 (21.74)	91 (19.78)	118 (25.65)	37 (8.04)
Non-Hispanic Black	277 (63.82)	164 (37.79)	102 (23.50)	96 (22.12)	50 (11.52)	95 (21.89)	27 (6.22)
Cancer diagnosis							
Have cancer	260 (71.23)	190 (52.05)	74 (20.27)	76 (20.82)	79 (21.64)	117 (32.05)	50 (13.70)
No cancer	1315 (60.85)	1068 (49.42)	425 (19.67)	495 (22.91)	356 (16.47)	430 (19.90)	178 (8.24)

The statistical analysis also yielded significant differences across the different demographics (Table 5). Older patients (≥65 years old) were more likely to avoid using online medical records. They preferred speaking directly to their health care providers (OR 1.76, 95% CI 1.35-2.31, $P<.001$), were more concerned about privacy issues caused by online medical records (OR 1.79, 95% CI 1.22-2.66, $P<.001$), and felt uncomfortable using the systems (OR 10.55, 95% CI 6.06-19.89, $P<.001$) compared to patients aged 18-34 years; however, they were more likely to be in need of online records (OR 0.47, 95% CI 0.36-0.62, $P<.001$).

Table 5. Reasons for not using online medical records.

Characteristic	Odds ratio (95% CI)						
	SpeakDirectly	NoNeed	ConcernedPrivacy	NoRecord	LogInProb	Uncomfortable	MultipleRec
Gender							
Male	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)
Female	1.13 (0.96-1.33)	0.81 (0.69-0.96)*	1.09 (0.89-1.34)	0.74 (0.61-0.90)**	1.21 (0.96-1.54)	0.92 (0.76-1.12)	0.87 (0.65-1.15)
Age (years)							
18-34	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)
35-49	0.92 (0.68-1.24)	0.76 (0.56-1.04)	1.57 (1.02-2.44)*	0.79 (0.56-1.11)	1.10 (0.72-1.70)	2.53 (1.34-5.06)**	0.92 (0.52-1.63)
50-64	1.73 (1.31-2.2)***	0.56 (0.42-0.74)***	2.01 (1.37-3.01)	0.65 (0.47-0.88)**	1.15 (0.78-1.70)	5.44 (3.08-10.36)*	1.17 (0.72-1.94)
>65	1.76 (1.35-2.31)***	0.47 (0.36-0.62)***	1.79 (1.22-2.66)**	0.57 (0.42-0.77)*	1.43 (0.99-2.09)	10.55 (6.06-19.89)***	1.12 (0.70-1.84)
Race							
Non-Hispanic White	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)
Hispanic	1.10 (0.88-1.38)	0.46 (0.36-0.57)***	1.42 (1.09-1.84)	0.93 (0.71-1.20)	1.12 (0.85-1.46)	1.34 (1.04-1.71)*	0.78 (0.52-1.14)
Non-Hispanic Black	1.10 (0.88-1.38)	0.46 (0.37-0.58)***	1.42 (1.09-1.84)	0.95 (0.72-1.23)	0.59 (0.42-0.82)**	1.08 (0.83-1.41)	0.59 (0.37-0.91)*
Cancer							
Have cancer	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	1.00 (1.00-1.00)
No cancer	0.62 (0.48-0.80)***	0.90 (0.71-1.13)	0.96 (0.72-1.28)	1.12 (0.85-1.50)	0.71 (0.53-0.95)*	0.52 (0.41-0.67)***	0.56 (0.40-0.80)**

*P<.05; **P<.01; ***P<.001.

Female respondents were more likely to use online medical records, and they were more likely to have records (OR 0.74, 95% CI 0.61-0.90, P=.003 for female). Patients who were Black or Hispanic were more concerned about privacy issues (Black: OR 1.42, 95% CI 1.09-1.84, P=.007; Hispanic: OR 1.42, 95% CI 1.09-1.84, P=.007) but were more conscious about the need for online records (Black: OR 0.46, 95% CI 0.37-0.58, P<.001; Hispanic: OR 0.46, 95% CI 0.36-0.57, P<.001). Patients who were Black were more likely to have problems logging in to their records (OR 0.59, 95% CI 0.42-0.82, P=.002).

Patients with cancer preferred to speak directly to the health care providers (OR 0.62, 95% CI 0.48-0.80, P<.001) and were more likely to feel uncomfortable using online medical records (OR 0.52, 95% CI 0.41-0.67, P<.001).

Discussion

General

Many studies that aim to improve information sharing and technology use in health care settings are based on exploring design improvements in patient-centered tools. Some discuss the environmental and technical barriers of adopting these tools [12,18], and others focus on providing training to enhance use-related skills [24]. We used nationally representative data from 2019 and examined attitude and other factors influencing nonuse of online medical records across different groups of patients. The results portray trends just prior to the pandemic, which has undoubtedly precipitated a paradigm shift toward online and telehealth medical use since the beginning of 2020.

The results showed that online medical record use was improving compared to use in previous years—41% overall in 2019 compared to 28% in 2017 [25]. However, online medical record use was still only approximately 30% for older adults (65 years and above) and respondents who were Black or Hispanic. Older adults were less likely to use online medical records compared to younger patients. Previous studies have also found that those older than 65 years would be less likely to use the internet to find health information [26,27] and less likely to use electronic personal health records [28]. Another study [29,30] also demonstrated that older adults have rather negative attitudes toward computers. Patients who were Black or Hispanic in our sample were more likely to be nonusers of online medical records than patients who were White. This was consistent with the findings of a previous study [27] that collected data between 2010 and 2017; however, in our study, there were increased rates of online medical record use among

minorities compared to those of previous years. Due to chronic underlying health conditions, adults aged ≥65 years and individuals who are Black or Hispanic were the two groups hit hardest by the COVID-19 pandemic [31]. These groups often need continued care due to chronic health conditions [32]; however, visiting hospitals and clinics during the pandemic may increase the risk of infection. The online medical use rate would likely to be increased among these groups due to the pandemic.

In this study, we explicitly focused on the reasons people avoid online medical record use across demographics. The major reason that emerged from the data is that patients would prefer to "see their physicians in person." This was the primary reason for all groups except younger patients. Older patients were more likely to prefer to "speak in person with physicians" than younger patients. Furthermore, the preference to "speak in person with physicians" was also high for all race groups (1575/2526, 62.35%). These findings are not surprising since most patients value in-person visits with their doctors due to the more personal nature of the interaction, the opportunity to use nonverbal cues, and the ability to explain specific symptoms to doctors more clearly. Since March of 2020, however, the health care system is experiencing a paradigm shift due to an unprecedented pandemic, and telehealth visits have become a new normal, often replacing in-person visits, especially for older and chronically ill patients who are at high risk for COVID-19 infection. There is a tremendous rise in virtual care via telehealth technology during the pandemic [33]. Future studies will undoubtedly show changes in the demand for in-person visits after these forced experiences.

In the event of a population-wide infectious disease outbreak such as COVID-19, people's online activities may affect public concerns and health behaviors. Many studies [34,35] have explored people's active use of online information in various crises, including a public health crisis. The recent COVID-19 pandemic also sparked a paradigm shift in using online health care communication tools such as telehealth. It showed the importance and necessity of information sharing and communication beyond the walls of clinics.

The technology acceptance model explains that people use technologies when they are perceived as useful and necessary [36]. Almost half of the nonusers stated that there was no need for them to use online medical records. Young respondents (aged 18-34 years) among the sample were shown to be more likely to use online electronic medical records. Those who did not use online medical records in this group were likely to express that they had "no need" for their online medical record during this time period. This showed a significant difference compared to older patients ($P<.001$). Intuitively, younger patients are healthier than older patients, but they also tend to be more comfortable using online health technologies. After the COVID-19 pandemic, a new necessity may emerge which also might influence young patients use of online medical records for any information exchange with their providers. Future HINTS data collected during or after the pandemic might also show this shift among young patients. Finally, respondents who were Black or Hispanic were less likely to state the reason of "no need" than respondents who were White.

Privacy has always been an issue for some users regarding the use of technologies for information sharing, especially information as sensitive and personal as medical records. Some participants also declared this as one of the factors for avoiding online medical record use. Respondents who were Black or Hispanic were more likely to have privacy concerns compared to those who were White. Historically, minorities have less trust in the health care system due to disparities they have experienced [24,37-39]. This might also influence their perception of privacy regarding any online health information exchange. Furthermore, older patients are highly likely to state "being uncomfortable using online record" as a reason for avoiding online medical records compared to younger patients. The US population had almost 52 million people older than 65 years as of 2018 [40]. This population will be the major consumer of health care systems for the foreseeable future; therefore, any online tools need to be redesigned to be user friendly (ie, for this population to use easily and comfortably).

Finally, our study also showed that patients with cancer use online medical records more than patients who do not have cancer. This was consistent with the findings of previous studies [41-43] showing that patients with such conditions may have a greater need for health tracking and sharing health information with multiple health care professionals than others. Among nonusers of online medical records, patients with cancer are likely to prefer speaking with physicians in person compared to patients who do not have cancer. The complexity of treatment options and the emotional aspect of visits make it more necessary for patients with cancer to meet with doctors in person. On the other hand, the health care system should have alternative plans to maintain quality online visits with these patients during the pandemic.

Limitations

This study also has limitations. First, the nature of HINTS data is cross-sectional and relies on subjective responses; therefore, it is not able to offer information on causality. Second, the low response rate (20%-30%) might raise some bias concerns, especially related to nonrespondents and sampling strategy. We should also note that the sampling and weighting strategy used by HINTS administrators helps minimize biases and improve national representativeness and generalizability of findings. Nonetheless, some local studies with more detail and a higher response rate should be conducted to validate the findings.

Conclusion

This study showed factors that lead people to avoid online medical record use across different demographics using a nationally representative survey. The findings show that there is an increased rate of online medical record use compared to previous years; however, this rate is still not at the expected level. The study shows that most patients still prefer speaking in person with their providers instead of using online medical records. Future studies should also look at how the education level of patients impacts these studied factors; our data did not have that component.

We also acknowledge that the recent COVID-19 pandemic has shifted the culture of virtual visits and online medical record

use in health care. Future studies should look at online medical record use trends and factors during and after the pandemic to see how these have shifted. Finally, future designs and concepts of online medical communication technologies may also consider the importance of preparing a common ground for patients where different technology acceptance levels are respected.

Acknowledgments

Research reported in this publication was supported by the National Institute of Nursing Research of the National Institutes of Health (award number R15NR018965). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Norouzinia R, Aghabarari M, Shiri M, Karimi M, Samami E. Communication barriers perceived by nurses and patients. Glob J Health Sci 2015 Sep 28;8(6):65-74 [FREE Full text] [doi: 10.5539/gjhs.v8n6p65] [Medline: 26755475]
2. Scaioli G, Schäfer WLA, Boerma WGW, Spreeuwenberg P, van den Berg M, Schellevis FG, et al. Patients' perception of communication at the interface between primary and secondary care: a cross-sectional survey in 34 countries. BMC Health Serv Res 2019 Dec 30;19(1):1018 [FREE Full text] [doi: 10.1186/s12913-019-4848-9] [Medline: 31888614]
3. Bartlett G, Blais R, Tamblyn R, Clermont RJ, MacGibbon B. Impact of patient communication problems on the risk of preventable adverse events in acute care settings. CMAJ 2008 Jun 03;178(12):1555-1562 [FREE Full text] [doi: 10.1503/cmaj.070690] [Medline: 18519903]
4. van Walraven C, Taljaard M, Bell CM, Etchells E, Zarnke KB, Stiell IG, et al. Information exchange among physicians caring for the same patient in the community. CMAJ 2008 Nov 04;179(10):1013-1018 [FREE Full text] [doi: 10.1503/cmaj.080430] [Medline: 18981442]
5. de Lusignan S, Teasdale S, Little D, Zapp J, Zuckerman A, Bates D, et al. Comprehensive computerised primary care records are an essential component of any national health information strategy: report from an international consensus conference. Inform Prim Care 2004 Dec 01;12(4):255-264 [FREE Full text] [doi: 10.14236/jhi.v12i4.133] [Medline: 15808027]
6. Mukai TO, Bro F, Fenger-Grøn M, Olesen F, Vedsted P. Use of hyperlinks in electronic test result communication: a survey study in general practice. BMC Med Inform Decis Mak 2012 Oct 04;12(1):114 [FREE Full text] [doi: 10.1186/1472-6947-12-114] [Medline: 23035761]
7. Wennberg JE. Time to tackle unwarranted variations in practice. BMJ 2011 Mar 17;342(mar17 3):d1513-d1513. [doi: 10.1136/bmj.d1513] [Medline: 21415111]
8. Elbourne D, Richardson M, Chalmers I, Waterhouse I, Holt E. The Newbury Maternity Care Study: a randomized controlled trial to assess a policy of women holding their own obstetric records. Br J Obstet Gynaecol 1987 Jul;94(7):612-619. [doi: 10.1111/j.1471-0528.1987.tb03165.x] [Medline: 3304403]
9. Latimer EJ, Porterfield P, Roberts JG. A patient care travelling record for palliative care: a feasibility study. J Palliat Care 2019 Apr 08;7(2):30-36. [doi: 10.1177/082585979100700206]
10. Shah P, Selwyn B, Shah K, Kumar V. Evaluation of the home-based maternal record: a WHO collaborative study. Bull World Health Organ 1993;71(5):535-548 [FREE Full text] [Medline: 8261557]
11. van Wersch A, de Boer M, van der Does E, de Jong P, Knegt P, Meeuwis C, et al. Continuity of information in cancer care: evaluation of a logbook. Patient Educ Couns 1997 Jul;31(3):223-236. [doi: 10.1016/s0738-3991(97)00030-x]
12. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
13. Krist AH, Woolf SH, Hochheimer C, Sabo RT, Kashiri P, Jones RM, et al. Harnessing information technology to inform patients facing routine decisions: cancer screening as a test case. Ann Fam Med 2017 May;15(3):217-224 [FREE Full text] [doi: 10.1370/afm.2063] [Medline: 28483886]
14. Lu H, Shaw BR, Gustafson DH. Online health consultation: examining uses of an interactive cancer communication tool by low-income women with breast cancer. Int J Med Inform 2011 Jul;80(7):518-528 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.03.011] [Medline: 21530381]
15. Otte-Trojel T, de Bont A, van de Klundert J, Rundall TG. Characteristics of patient portals developed in the context of health information exchanges: early policy effects of incentives in the meaningful use program in the United States. J Med Internet Res 2014 Nov 21;16(11):e258 [FREE Full text] [doi: 10.2196/jmir.3698] [Medline: 25447837]

16. Gerard M, Fossa A, Folcarelli PH, Walker J, Bell SK. What patients value about reading visit notes: a qualitative inquiry of patient experiences with their health information. J Med Internet Res 2017 Jul 14;19(7):e237 [FREE Full text] [doi: 10.2196/jmir.7212] [Medline: 28710055]
17. Black H, Gonzalez R, Priolo C, Schapira MM, Sonnad SS, Hanson CW, et al. True "meaningful use": technology meets both patient and provider needs. Am J Manag Care 2015 May 01;21(5):e329-e337 [FREE Full text] [Medline: 26167781]
18. de Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open 2014 Sep 08;4(9):e006021 [FREE Full text] [doi: 10.1136/bmjopen-2014-006021] [Medline: 25200561]
19. Patel V, Barker W, Siminerio E. Individuals' access and use of their online medical record nationwide. ONC Data Brief 2014(20):1-12.
20. Demiris G. Consumer health informatics: past, present, and future of a rapidly evolving domain. Yearb Med Inform 2016 May 20;Suppl 1:S42-S47 [FREE Full text] [doi: 10.15265/IYS-2016-s005] [Medline: 27199196]
21. Nelson DE, Kreps GL, Hesse BW, Croyle RT, Willis G, Arora NK, et al. The Health Information National Trends Survey (HINTS): development, design, and dissemination. J Health Commun 2004 Sep;9(5):443-60; discussion 81. [doi: 10.1080/10810730490504233] [Medline: 15513791]
22. Health Information National Trends Survey. National Cancer Institute. 2019. URL: http://hints.cancer.gov/ [accessed 2020-09-09]
23. Finney RLJ, Davis T, Beckjord EB, Blake K, Moser RP, Hesse BW. Picking up the pace: changes in method and frame for the health information national trends survey (2011-2014). J Health Commun 2012;17(8):979-989 [FREE Full text] [doi: 10.1080/10810730.2012.700998] [Medline: 23020763]
24. Gordon NP, Hornbrook MC. Differences in access to and preferences for using patient portals and other ehealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
25. Patel V, Johnson C. Individuals use of online medical records and technology for health needs. ONC Data Brief 2018(40):1-12.
26. Czaja SJ, Sharit J, Nair SN, Lee CC. Older adults and internet health information seeking. Proc Hum Factors Ergon Soc Annu Meet 2016 Nov 05;53(2):126-130. [doi: 10.1177/154193120905300204]
27. Asan O, Cooper II F, Nagavally S, Walker RJ, Williams JS, Ozieh MN, et al. Preferences for health information technologies among US adults: analysis of the Health Information National Trends Survey. J Med Internet Res 2018 Oct 18;20(10):e277 [FREE Full text] [doi: 10.2196/jmir.9436] [Medline: 30341048]
28. Wen K, Kreps G, Zhu F, Miller S. Consumers' perceptions about and use of the internet for personal health records and health information exchange: analysis of the 2007 Health Information National Trends Survey. J Med Internet Res 2010 Dec;12(4):e73 [FREE Full text] [doi: 10.2196/jmir.1668] [Medline: 21169163]
29. Wandke H, Sengpiel M, Sönksen M. Myths about older people's use of information and communication technology. Gerontology 2012;58(6):564-570 [FREE Full text] [doi: 10.1159/000339104] [Medline: 22739502]
30. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
31. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HHX, Mercer SW, et al. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. Lancet Glob Health 2020 Aug;8(8):e1003-e1017. [doi: 10.1016/S2214-109X(20)30264-3]
32. Khera A, Baum S, Gluckman T, Gulati M, Martin S, Michos E, et al. Continuity of care and outpatient management for patients with and at high risk for cardiovascular disease during the COVID-19 pandemic: A scientific statement from the American Society for Preventive Cardiology. Am J Prev Cardiol 2020 Mar;1:100009 [FREE Full text] [doi: 10.1016/j.ajpc.2020.100009] [Medline: 32835347]
33. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
34. Gui X, Kou Y, Pine KH, Chen Y. Managing uncertainty: using social media for risk assessment during a public health crisis. 2017 Presented at: Conference on Human Factors in Computing Systems; May 6-11; Denver, Colorado p. 4520-4533. [doi: 10.1145/3025453.3025891]
35. Rizo CA, Lupea D, Baybourdy H, Anderson M, Closson T, Jadad AR. What internet services would patients like from hospitals during an epidemic? Lessons from the SARS outbreak in Toronto. J Med Internet Res 2005 Aug 03;7(4):e46 [FREE Full text] [doi: 10.2196/jmir.7.4.e46] [Medline: 16236698]
36. Holden RJ, Asan O, Wozniak EM, Flynn KE, Scanlon MC. Nurses' perceptions, acceptance, and use of a novel in-room pediatric ICU technology: testing an expanded technology acceptance model. BMC Med Inform Decis Mak 2016 Dec 15;16(1):145 [FREE Full text] [doi: 10.1186/s12911-016-0388-y] [Medline: 27846827]
37. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 2011 Oct;26(10):1112-1116 [FREE Full text] [doi: 10.1007/s11606-011-1728-3] [Medline: 21538166]

38. Smith SG, O'Conor R, Aitken W, Curtis LM, Wolf MS, Goel MS. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015 Jul;22(4):888-895 [FREE Full text] [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
39. Yamin CK, Emani S, Williams DH, Lipsitz SR, Karson AS, Wald JS, et al. The digital divide in adoption and use of a personal health record. Arch Intern Med 2011 Mar 28;171(6):568-574. [doi: 10.1001/archinternmed.2011.34] [Medline: 21444847]
40. Mather M, Scommegna P. Fact sheet: aging in the United States. Population Reference Bureau. URL: https://www.prb.org/aging-unitedstates-fact-sheet/ [accessed 2020-10-10]
41. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 1;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
42. Kuwabara A, Su S, Krauss J. Utilizing digital health technologies for patient education in lifestyle medicine. Am J Lifestyle Med 2020 Dec 13;14(2):137-142 [FREE Full text] [doi: 10.1177/1559827619892547] [Medline: 32231478]
43. Baker JM, Grant RW, Gopalan A. A systematic review of care management interventions targeting multimorbidity and high care utilization. BMC Health Serv Res 2018 Jan 30;18(1):65 [FREE Full text] [doi: 10.1186/s12913-018-2881-8] [Medline: 29382327]

Abbreviations

HINTS: Health Information National Trends Survey

Edited by G Eysenbach; submitted 04.10.20; peer-reviewed by M Rigby, M Reblin; comments to author 12.11.20; revised version received 19.11.20; accepted 12.12.20; published 22.02.21.

Please cite as:
Elkefi S, Yu Z, Asan O
Online Medical Record Nonuse Among Patients: Data Analysis Study of the 2019 Health Information National Trends Survey
J Med Internet Res 2021;23(2):e24767
URL: https://www.jmir.org/2021/2/e24767
doi:10.2196/24767
PMID:33616539

Review

Quantifying Patient Portal Use: Systematic Review of Utilization Metrics

Lauren L Beal[1,2], BA; Jacob M Kolman[1], MA; Stephen L Jones[1,3], MSHI, MD; Aroub Khleif[4], PhD; Terri Menser[1,3], MBA, PhD

[1]Center for Outcomes Research, Houston Methodist Hospital, Houston, TX, United States

[2]University of Texas Health Science Center, McGovern Medical School, Houston, TX, United States

[3]Department of Surgery, Weill Cornell Medical College, New York, NY, United States

[4]Ambulatory Clinical Systems, Information Technology Division, Houston Methodist Hospital, Houston, TX, United States

Corresponding Author:
Terri Menser, MBA, PhD
Center for Outcomes Research, Houston Methodist Hospital
7550 Greenbriar Dr
Suite JRB4-100
Houston, TX, 77030
United States
Phone: 1 270 625 2749
Email: tmenser@houstonmethodist.org

Abstract

Background: Use of patient portals has been associated with positive outcomes in patient engagement and satisfaction. Portal studies have also connected portal use, as well as the nature of users' interactions with portals, and the contents of their generated data to meaningful cost and quality outcomes. Incentive programs in the United States have encouraged uptake of health information technology, including patient portals, by setting standards for meaningful use of such technology. However, despite widespread interest in patient portal use and adoption, studies on patient portals differ in actual metrics used to operationalize and track utilization, leading to unsystematic and incommensurable characterizations of use. No known review has systematically assessed the measurements used to investigate patient portal utilization.

Objective: The objective of this study was to apply systematic review criteria to identify and compare methods for quantifying and reporting patient portal use.

Methods: Original studies with quantifiable metrics of portal use published in English between 2014 and the search date of October 17, 2018, were obtained from PubMed using the Medical Subject Heading term "Patient Portals" and related keyword searches. The first search round included full text review of all results to confirm a priori data charting elements of interest and suggest additional categories inductively; this round was supplemented by the retrieval of works cited in systematic reviews (based on title screening of all citations). An additional search round included broader keywords identified during the full-text review of the first round. Second round results were screened at abstract level for inclusion and confirmed by at least two raters. Included studies were analyzed for metrics related to basic use/adoption, frequency of use, duration metrics, intensity of use, and stratification of users into "super user" or high utilizers. Additional categories related to provider (including care team/administrative) use of the portal were identified inductively. Additional analyses included metrics aligned with meaningful use stage 2 (MU-2) categories employed by the US Centers for Medicare and Medicaid Services and the association between the number of portal metrics examined and the number of citations and the journal impact factor.

Results: Of 315 distinct search results, 87 met the inclusion criteria. Of the a priori metrics, plus provider use, most studies included either three (26 studies, 30%) or four (23 studies, 26%) metrics. Nine studies (10%) only reported the patient use/adoption metric and only one study (1%) reported all six metrics. Of the US-based studies (n=76), 18 (24%) were explicitly motivated by MU-2 compliance; 40 studies (53%) at least mentioned these incentives, but only 6 studies (8%) presented metrics from which compliance rates could be inferred. Finally, the number of metrics examined was not associated with either the number of citations or the publishing journal's impact factor.

Conclusions: Portal utilization measures in the research literature can fall below established standards for "meaningful" or they can substantively exceed those standards in the type and number of utilization properties measured. Understanding how patient

portal use has been defined and operationalized may encourage more consistent, well-defined, and perhaps more meaningful standards for utilization, informing future portal development.

(J Med Internet Res 2021;23(2):e23493) doi:10.2196/23493

KEYWORDS

patient portals; meaningful use; American Recovery and Reinvestment Act; Health Information Technology for Economic and Clinical Health Act; portal utilization; patient-generated health data; portal; systematic review

Introduction

A patient portal is a secure online website, managed by a health care organization, that provides patients access to their personal health information [1-3]. Portals were developed to provide patients with a platform through which to claim ownership over their health care. For patients that adopt health care portals, usage of the portal has been shown to positively impact health outcomes [1]. Despite their introduction in the late 1990s to augment patient engagement [2], widespread adoption of patient portals was not seen until 2006 [2,4]. As of 2018, a reported 90% of health care organizations offer patients portal access, with the remaining 10% reporting plans to adopt this tool [5].

Numerous studies have investigated the relationship between patient portal utilization and health outcomes, specifically indicating a link between increased portal use and increased rates of patient engagement [6-9]. Notably, engaged individuals more actively participate in the management of their health care [10] and report enhanced patient satisfaction [11], a finding increasingly critical in patients with chronic diseases [12]. Patient portal utilization has been linked to "significant decreases in office visits…, changes in medication regimen, and better adherence to treatment" [13], along with improved chronic disease management and disease awareness [8,9]. Interestingly, even the content of patient messages was recently found to be associated with estimated readmission rates in patients with ischemic heart disease [14]. In these ways, patient portals have been cited as essential components of the solution to the cost and quality health care crisis in the United States [2].

A driving force behind the adoption and current progression of patient portals is the meaningful use (MU) criteria [13,15,16]. Introduced in 2009, the American Recovery and Reinvestment Act [2,16] included $30 billion [17] for the incentive program's implementation to fund government reimbursements for patient-centered health care [13] with the goal of utilizing electronic exchange of health information to improve quality of care [2]. Specific program guidelines, including an emphasis on increasing patient-controlled data and financial incentives to interact with patients through a patient portal [1,18], resulted in increased portal utilization [1]. Interactive, MU-mandated features of patient portals currently include (1) a clinical summary following each patient visit, (2) support of secure messaging between the patient and health care provider, and (3) the functionality of viewing, downloading, and transferring patient data [2].

Coupled with advances in technology and continued movement toward focusing on patient-centered care, features beyond those described by MU criteria have been implemented, including online appointment scheduling and bill payments, and continue to shape portal evolution. Mirroring the benefits of this technology, numerical projections demonstrate that the rate at which patients wish to utilize patient portals far exceeds the rate at which this technology is provided to them by their health care providers [19], with an estimated 75% of individuals accessing their personal health records via patient portals by 2020 [19].

Despite widespread portal interest and adoption, as well as comprehensive reviews on patient engagement with portals [2], no review has systematically assessed measurements investigating patient portal utilization. Currently, measurement of patient portal use varies widely, with inconsistent conceptual definitions serving as a consistent limitation to robust analysis [20]. Understanding how patient portal use has been defined and operationalized, both previously and currently, will encourage consistent and well-defined utilization of patient portals. Further, standardization of patient portal measurements will provide a basis from which to systematically analyze how to continue developing patient portals best suited to consumer needs.

Methods

Study Eligibility Criteria

This systematic review includes original studies with quantifiable metrics of portal use, broadly construed. Subjective reporting on usability, design requirements, or other qualitative analyses were excluded as nontopical. Systematic reviews were also excluded, although their bibliographies were utilized for reference crawling. The criteria used to determine eligibility of studies employing self-reported use and prospective studies emerged inductively through interrater review and discussion (between TM, LLB, and JMK) based on preliminary results. Self-reporting measures were excluded unless they reported direct portal usage data that were quantifiable and similar to actual portal use tracking (eg, by frequency of logins, duration of sessions, number of functions used, etc). Prospective trial designs were omitted if they artificially influenced portal use but were included if the portal use metric could be reasonably abstracted from its experimental context (eg, as either a quantified outcome or an uncontrolled baseline measure).

Studies available in English, published between 2014 and the end search date of October 17, 2018, were eligible for inclusion. The year 2014 was selected due to the full rollout of Centers for Medicare & Medicaid Services (CMS) meaningful use (MU) stage 2 (MU-2) requirements and the emergence of "Patient Portals" as a Medical Subject Headings (MeSH) term (automatically including cognate terms "Patient Internet Portal,"

"Patient Portal," "Patient Web Portal," and "Patient Web Portals," and subsumed under "Health Records, Personal," with previous indexing via the less focused term, "Electronic Health Records," from 2010-2016). Although US MU-2 regulations were of particular interest, we did not exclude studies from other countries so that potentially informative use metrics employed outside the United States would also inform the results. Studies from outside the United States were not included in the analysis focused on MU.

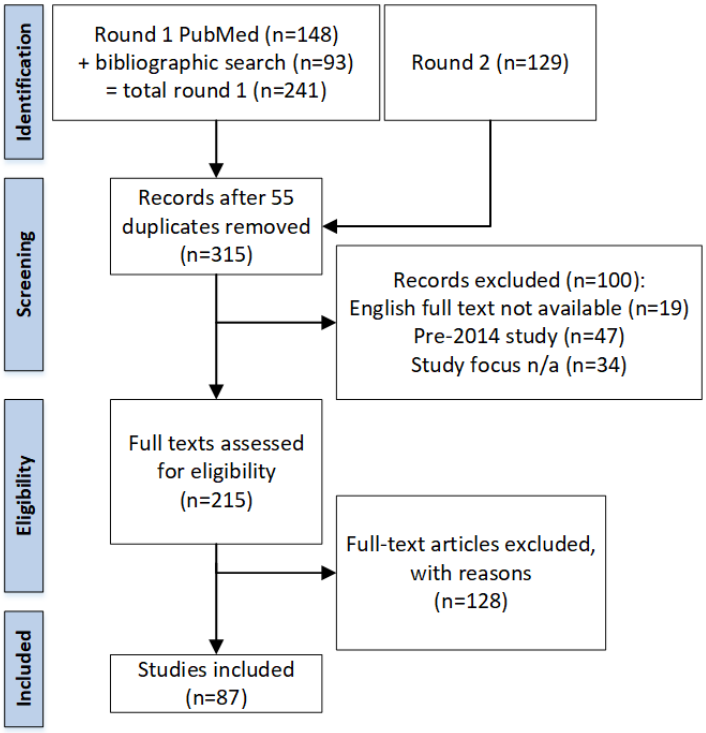
Identification and Selection of Studies

The search proceeded in two major rounds. In round 1, authors JMK and LLB identified studies in PubMed by applying the MeSH term "Patient Portals" with no other limiters. All round 1 results were initially reviewed in English full text if available rather than relying on title/abstract screening alone. Full texts were read to better orient the raters to the patient portal utilization literature, suggest secondary sources and search terms, confirm a priori data charting categories of interest (use/adoption, frequency, duration, intensity, and super user), and suggest additional charting elements inductively. Within the round 1 results, secondary searching was performed on all article bibliographies, and relevant titles were retrieved for review. Results prior to 2014 were excluded based on the emerging relevance and centrality of CMS MU.

After reviewing the literature (241 full-text articles from round 1, including 148 articles from PubMed and 93 articles from the bibliographic search), themes that emerged as commonly studies metrics became the basis of our coding for the complete two-part search process. These confirmed our a priori themes of interest and would serve as relevant limiters in the broader keyword-based second search to complement the more restrictive round 1 MeSH-based search. Round 2 searching (by TM) applied the following terms at title/abstract and keyword levels: "patient portal" AND (frequency OR use OR duration OR intensity). Duplicate results were removed by an automatic process using Stata matching on title, author, and year, followed by a manual check to remove additional duplicates that were missed (eg, formatting, punctuation, etc).

The full results were screened at the abstract level to exclude non–English articles, those without full text, sources older than 2014, and articles lacking quantified portal metrics. Two raters (LLB and JMK) assessed inclusion at the abstract level, followed by full text review; any inclusion or data charting discrepancies were resolved in rater meetings (by LLB, JMK, and TM). Figure 1 summarizes the screening and inclusion process.

Figure 1. Study selection and inclusion process reported per guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [21]. n/a: Not applicable.



Data Collection and Study Appraisal

For coding purposes, use/adoption, frequency, duration, intensity, and super user (or similar user stratification) were considered a priori themes from which to extract definitions; provider use emerged as a theme inductively. Super user, in this context, is synonymous with high utilizer and should not be confused with the information technology standard definition implying a user with elevated privileges. All metrics were coded as binary, indicating the presence of a measure for and/or definition of each respective metric. These data were coded and recorded in a spreadsheet containing the article citation information and columns for themes of interest for both portal use metric definitions and MU criteria. Extractors' working definitions of metric types are summarized in Table 1.

Table 1. Study inclusion metric definitions.

Metric	Definition used for data charting
Provider use	• Portal use by providers, care teams, or other staff. This use could be in terms of adoption, frequency, intensity, duration, or super user, per below; patient utilization grouped by provider practice/specialty also implies provider/practice adoption.
Patient use/adoption	• Patient use/adoption, including proxies, parents, or surrogates acting on behalf of patients. Total number/percentage of registrations, logins, or other basic access information falls under this category.
Frequency	• Metrics that involve count variables per time unit or within a study window (eg, patient logins per week, message rates per month, functions used during an inpatient window).
Duration	• Duration as a continuous time variable (eg, an average patient login session time of "n" minutes, months/years during which a user account remained active, time from visit to first login); "length" variables for message threads (two vs three messages per thread) were also interpreted as a type of duration, although only in one instance.
Intensity	• Counts that include a qualitative "depth" component beyond clicks or page accesses, including counts of messages with clinical relevance, message threads leading to clinical resolution, and use of functions entailing substantive input/data entry, such as diaries or care preference plans.
Super user	• Users stratified in a way that distinguishes a high-utilization or high-activity group (eg, in terms of greater intensity, a categorically higher frequency, consistent duration of use, etc).

For included articles from the United States, a set of CMS MU themes emerged as relevant and were added upon agreement by all raters. An "MU motive" coding tag indicated either explicit desire for MU compliance in the article or at least mention of MU (eg, in the introduction or discussion). "MU consistent" tags rated apparent compatibility between the metric used and basic MU-2 requirements (regardless of motive), and included (1) "access"—coded if the percentage of patients accessing the portal information could be derived; (2) "send"—coded if patient-initiated messaging was tracked; (3) "view-download-transmit" (VDT)—coded if the percentage of patients who could VDT (specifically within the 4-day window following an office visit or 36 hours after discharge from hospital) could be determined from the metrics reported; and (4) MU-2—coded as shorthand when all three conditions, which together entail the full MU-2 compliance requirements, were met.

Finally, in lieu of assessing the methodological quality of these wide-ranging patient portal studies, we assessed quality based on two criteria: the quality of the journals in which articles were published based on their 2019 impact factor (except in one instance where the 2018 impact factor was used), and the citation count of each article. For the latter, we extracted the total number of citations (as of September 22, 2020) to calculate the mean number of citations per year that each article received based on the time period that elapsed since online publication date. Regression analyses were conducted to determine if the number of use metrics was predictive of the mean number of citations per year or the impact factor of the journal in which the article was published.

Results

A total of 315 search results remained after the removal of duplicates. All 315 articles were examined for defined patient portal metrics, with records excluded (n=100) for the following reasons: lack of the full-text English-language article or a suitably detailed abstract (non–English-language [n=18] or no text available [n=1]), study publication date prior to 2014 (n=47), and/or nonapplicable study focus (n=34). The remaining 215 studies were analyzed; of these, 128 were excluded, leaving 87 studies for inclusion in the analysis (see Figure 1). Notably, the abstracts (or translations thereof) of 18 non–English-language exclusions also met other exclusion criteria (eg, qualitative only/no metrics defined, portal development or usability studies, or literature reviews of unrelated portal topics).

Patient use was the most commonly studied patient portal metric, analyzed in 90% (78/87) of studies. Super user designations were only found in 24% (21/87) of studies, making this the least commonly studied metric. Table 2 identifies the frequency with which each metric was included in each study, with totals for each metric [6-10,18,22-102]. There were 32 different combinations of study metrics, identified in Table 3, with the two most common metric combinations being patient use/adoption, frequency, and intensity (n=9) and patient use/adoption alone (n=9). The majority of studies (53/87, 61%) analyzed three or fewer metrics, with 3.11 as the average number of metrics reported. The definitions of these 271 metrics are summarized by study in Multimedia Appendix 1.

Table 2. Frequencies of metric inclusion in analyzed studies (N=87).

	Metric[a]					
Study	Provider use	Patient use/adoption	Frequency	Duration	Intensity	Super user
Ackerman et al, 2017 [60]	✓[b]	✓	✓	✓	✓	
Ahmedani et al, 2016 [92]		✓	✓			
Aljabri et al, 2018 [69]		✓			✓	
Alpert et al, 2017 [98]	✓		✓	✓	✓	✓
Arcury et al, 2017 [47]		✓	✓		✓	
Bajracharya et al, 2016 [89]		✓	✓		✓	
Baldwin et al, 2017 [83]		✓				
Bell et al, 2018 [50]		✓	✓		✓	
Boogerd et al, 2017 [31]	✓	✓[c]	✓			✓
Bose-Brill et al, 2018 [67]	✓		✓		✓	✓
Bower et al, 2017 [93]		✓				
Chung et al, 2017 [70]	✓	✓[c]	✓		✓	✓
Crotty et al, 2014 [58]	✓	✓	✓		✓	
Crotty et al, 2015 [100]	✓		✓	✓	✓	
Dalal et al, 2016 [101]	✓	✓[c]	✓		✓	
Davis et al, 2015 [96]		✓			✓	
Devkota et al, 2016 [51]		✓			✓	
Dexter et al, 2016 [77]		✓				
Emani et al, 2016 [30]		✓	✓	✓	✓	
Fiks et al, 2015 [68]			✓			✓
Fiks et al, 2016 [53]		✓			✓	
Forster et al, 2015 [22]		✓	✓	✓	✓	
Garrido et al, 2014 [39]	✓	✓	✓		✓	✓
Garrido et al, 2015 [23]		✓				
Gheorghiu and Hagen, 2017 [24]		✓	✓			
Gordon and Hornbrook, 2016 [62]		✓			✓	
Graetz et al, 2016 [86]		✓	✓			
Griffin et al, 2016 [18]		✓	✓		✓	✓
Groen et al, 2017 [49]		✓	✓	✓		
Haun et al, 2014 [73]	✓	✓			✓	
Henry et al, 2016 [46]		✓				
Jhamb et al, 2015 [44]		✓	✓		✓	
Jones et al, 2015 [71]		✓	✓	✓	✓	✓
Kamo et al, 2017 [56]		✓	✓	✓	✓	
Kelly et al, 2017 [72]		✓	✓		✓	✓
King et al, 2017 [34]	✓	✓	✓	✓	✓	
Kipping et al, 2016 [25]		✓			✓	
Krasowski et al, 2017 [36]	✓	✓		✓		
Krist et al, 2014 [61]		✓	✓	✓		
Laccetti et al, 2016 [41]	✓	✓	✓		✓	✓

	Metric[a]					
Study	Provider use	Patient use/adoption	Frequency	Duration	Intensity	Super user
Lau et al, 2014 [29]		✓				
Lyles et al, 2016 [9]		✓[c]	✓		✓	✓
Mafi et al, 2016 [45]		✓	✓	✓		
Manard et al, 2016 [35]		✓			✓	
Masterman et al, 2017 [99]	✓				✓	
Masterson Creber et al, 2016 [7]			✓	✓	✓	
Mickles and Mielenz, 2014 [78]	✓	✓	✓	✓	✓	
Miles et al, 2016 [82]		✓[c]				
Mook et al, 2018 [74]		✓				
Neuner et al, 2015 [42]	✓	✓	✓	✓	✓	✓
North et al, 2014 [102]		✓[c]	✓	✓		✓
Oest et al, 2018 [80]		✓			✓	
Payne et al, 2016 [84]	✓	✓				
Pearl, 2014 [38]	✓	✓		✓	✓	
Pecina et al, 2017 [54]	✓	✓[c]	✓			
Peremislov, 2017 [91]	✓	✓[c]		✓	✓	
Perzynski et al, 2017 [63]		✓	✓		✓	
Petullo et al, 2016 [75]		✓	✓		✓	
Phelps et al, 2014 [52]		✓	✓	✓		✓
Pillemer et al, 2016 [37]	✓	✓			✓	
Price-Haywood and Luo, 2017 [6]		✓				
Quinn et al, 2018 [97]		✓[c]	✓		✓	
Redelmeier and Kraus, 2018 [26]		✓	✓		✓	
Reed et al, 2015 [57]	✓	✓	✓			
Reicher and Reicher, 2016 [64]		✓		✓	✓	
Riippa et al, 2014 [27]		✓	✓		✓	✓
Robinson et al, 2017 [81]		✓	✓		✓	✓
Ronda et al, 2014 [48]		✓	✓			
Runaas et al, 2017 [28]	✓		✓	✓		
Sarkar et al, 2014 [8]		✓[c]			✓	✓
Shaw et al, 2017 [65]	✓	✓		✓	✓	
Shenson et al, 2016 [76]	✓	✓	✓			
Shimada et al, 2016 [95]		✓[c]	✓	✓	✓	
Smith et al, 2015 [85]	✓	✓	✓		✓	
Sorondo et al, 2017 [43]		✓	✓		✓	
Steitz et al, 2017 [87]		✓	✓	✓	✓	
Thompson et al, 2016 [32]	✓	✓			✓	
Toscos et al, 2016 [40]			✓		✓	✓
Tulu et al, 2016 [10]	✓	✓			✓	
Vydra et al, 2015 [90]	✓	✓	✓	✓	✓	

	Metric[a]					
Study	Provider use	Patient use/adoption	Frequency	Duration	Intensity	Super user
Wallace et al, 2016 [88]		✓	✓		✓	✓
Weisner et al, 2016 [66]			✓		✓	
Williamson et al, 2017 [55]		✓	✓	✓	✓	
Wolcott et al, 2017 [59]	✓	✓[c]	✓		✓	✓
Wolff et al, 2016 [94]		✓			✓	
Woods et al, 2017 [33]		✓	✓			✓
Zhong et al, 2018 [79]		✓		✓		
Total	30 (34%)	78 (90%)	56 (64%)	27 (31%)	59 (68%)	21 (24%)

[a]See Multimedia Appendix 1 for full definition of each metric from each article.

[b]Indicates presence of the metric.

[c]Special contexts shaped the form of the metric in ways atypical of direct use analysis (eg, due to experimental controls).

MU serves as a driving criterion for patient portal adoption and utilization, reflected by the 87% (66/76) of US publications that included MU criteria, irrespective of an explicit motive, and 24% (18/76) of US studies explicitly implicating MU criteria as a driving force behind their publication. However, 33% (29/87) of the total manuscripts did not include any MU motive, whether because of foreign publication (n=11) or, among the 76 US sources, lack of statement of MU motivation (n=17) or a statement that the data and analysis did not fit the MU criteria because the study conduct predated its release (n=1). These studies indicate that investigation of metrics surrounding portal utilization extends beyond MU motives and metrics. Lau and colleagues [29] mentioned that while exact login time stamps were noted to be available, frequency, duration, and intensity metrics went unanalyzed in favor of a simplified metric. Similarly, the study conducted by Emani et al [30] mentioned MU stage 1 requirements, yet researchers purposefully extended the CMS 3-day window to 5 days for patients to access their postvisit summary. Further, Boogerd and colleagues [31] analyzed portal implementation through portal inaccessibility; this was achieved through measuring login difficulties and downtime while stratifying implementation efficacy through self-reported parenting stress, an efficacy measure not widely seen in the published literature. Mirroring Boogerd, Thompson et al [32] included information regarding the number of patient and proxy portal password changes, and Woods et al [33] reported on the number of unsuccessful or incomplete logins per study subject. Thus, several studies examined meaningful metrics of portal use and usability exceeding, or at least not anticipated by, MU-2 requirements. These metrics reflect portal utilization beyond the criteria of MU.

The patient use/adoption metric was the most frequently studied of the analyzed variables, included as a study metric in 90% (78/87) of studies (Table 2). Comparatively, provider use was analyzed in only 34% (30/87) of studies and rarely studied without simultaneously investigating patient use/adoption (5/87, 6%). Teasing apart patient use from provider use is an important distinction; however, some studies combine these distinct data points together and analyze the summed use [34]. Mirroring this, other studies group together patients not registered for the portal with registered patients that haven't messaged, highlighting the variability in reported metrics [35]. Similarly to the definitions of other analyzed metrics, provider use definitions revealed variability: while Krasowski et al [36] and Pillemer et al [37] tabulated provider use through "manual release of test results ahead of automatic release," others calculated this metric through provider response to patient messages [38,39]. The combined analysis of patient and provider use continues in more recent literature. Margolius et al [103] found that having a wealthier or larger patient population and working more days per week resulted in primary care physicians receiving more messages from patients, which the authors stratified into message types. Provider use has been shown to lead to patient use [20], but while patients can be led toward engagement with a system required by their physicians, they can also be led away from a system not utilized by their providers [104] (eg, if patients message their providers but don't receive a response). As mentioned previously, patient portal utilization has been employed as a proxy for patient engagement, with increased portal usage associated with better patient outcomes [105].

Notably, of the investigated metric groupings seen in Table 3, 59 studies (68%) included intensity as an analyzed metric, signaling the perceived importance of the depth at which patients were engaging with the portal. The definitions utilized by these studies varied: Emani and colleagues [30] distinguished between portal sessions and portal message use; Toscos et al [40] included intensity as the proportion of users engaging the daily health diary function; and Laccetti et al [41] investigated "staff MyChart actions performed per patient-initiated message."

Table 3. Number of metrics and metric combinations analyzed in 87 studies.

Number of metrics analyzed	Studies analyzing stated metrics, n (%)	Metric combinations
1	9 (10)	• Patient use/adoption (n=9)
2	18 (21)	• Patient use/adoption + intensity (n=9) • Patient use/adoption + frequency (n=4) • Patient use/adoption + duration (n=1) • Provider use + patient use/adoption (n=1) • Provider use + intensity (n=1) • Frequency + intensity (n=1) • Frequently + super user (n=1)
3	26 (30)	• Patient use/adoption + frequency + intensity (n=9) • Patient use/adoption + frequency + duration (n=3) • Patient use/adoption + provider use + intensity (n=4) • Patient use/adoption + provider use + frequency (n=3) • Frequency + intensity + super user (n=1) • Provider use + frequency + duration (n=1) • Frequency + duration + intensity (n=1) • Patient use/adoption + intensity + super user (n=1) • Provider use + patient use/adoption + duration (n=1) • Patient use/adoption + duration + intensity (n=1) • Patient use/adoption + frequency + super user (n=1)
4	23 (26)	• Patient use/adoption + frequency + intensity + super user (n=6) • Patient use/adoption + frequency + duration + intensity (n=6) • Provider use + patient use/adoption + duration + intensity (n=3) • Provider use + patient use/adoption + frequency + intensity (n=3) • Patient use/adoption + frequency + duration + super user (n=2) • Provider use + frequency + intensity + super user (n=1) • Patient use/adoption + provider use + frequency + super user (n=1) • Provider use + frequency + duration + intensity (n=1)
5	10 (11)	• Provider use + patient use/adoption + frequency + intensity + super user (n=4) • Provider use + patient use/adoption + frequency + duration + intensity (n=4) • Patient use/adoption + frequency + duration + intensity + super user (n=1) • Provider use + frequency + duration + intensity + super user (n=1)
6	1 (1)	• Provider use + patient use/adoption + frequency + duration + intensity + super user (n=1)

Neither of the quality variables (ie, journal impact factor and citation count) were shown to be statistically significantly associated with the number of patient portal metrics described in Table 3. The relationship between the number of patient portal metrics examined and the journal impact factor and the mean number of citations per year are visually depicted in Figures 2 and 3, respectively.

Figure 2. Relationship between the number of portal measures examined in an article and the impact factor of the publishing journal.

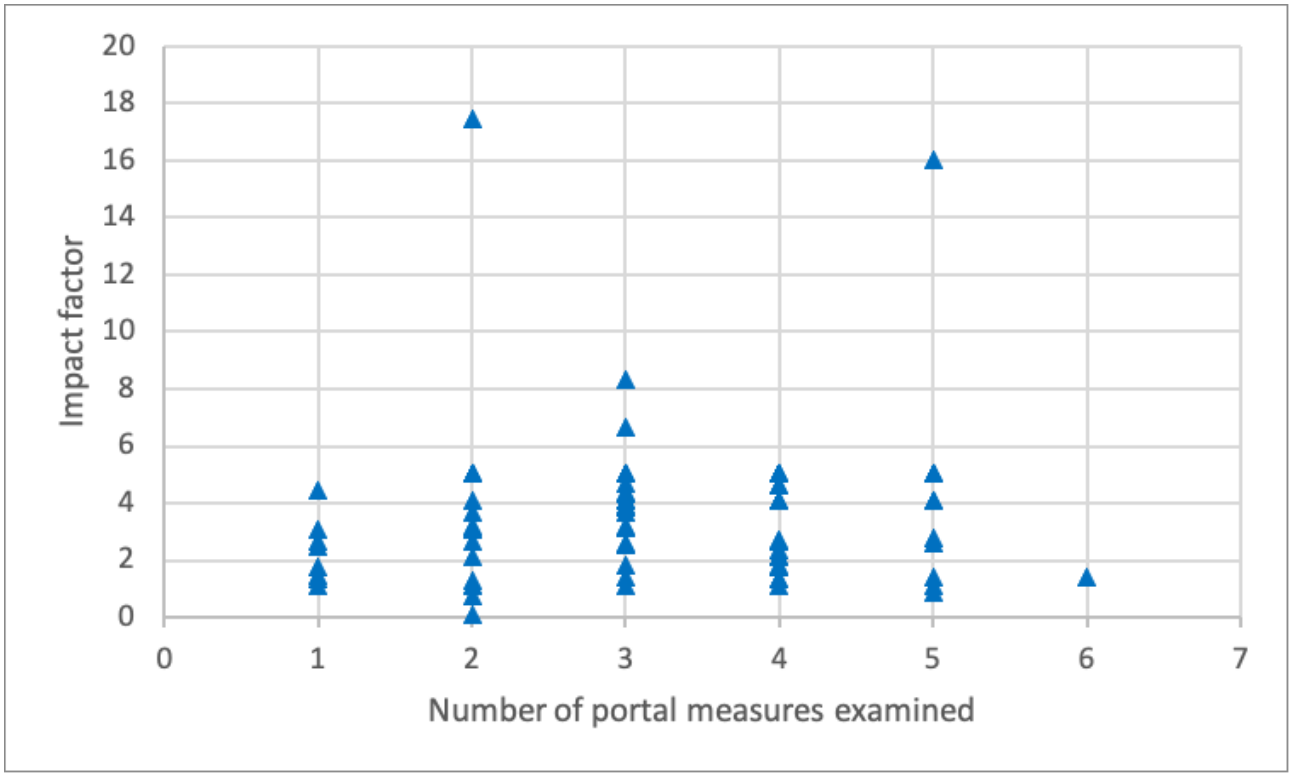


Figure 3. Relationship between the number of portal measures examined in an article and the mean number of citations per year (via Google Scholar).

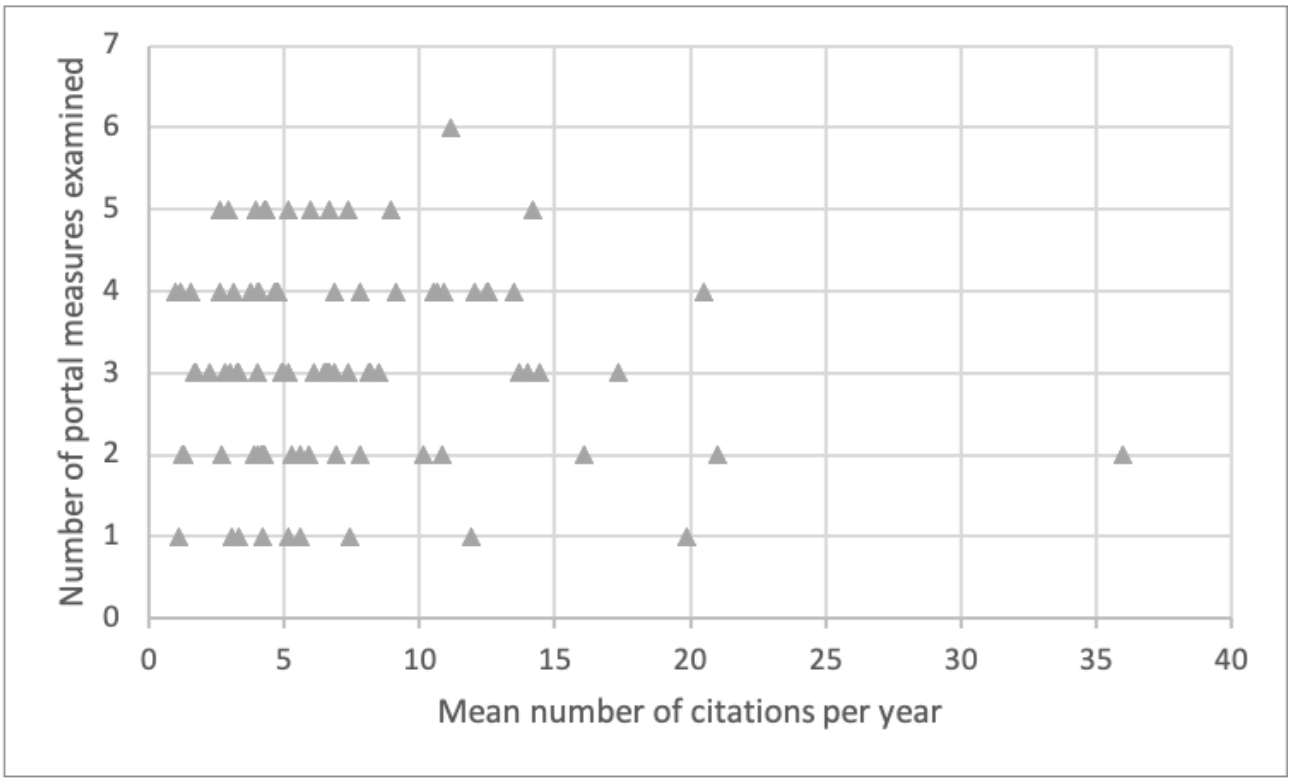


Table 4 depicts studies that were consistent with some metric of MU criteria. Articles not published in the United States were excluded from MU analysis due to variability in portal guidelines by country. Combining the three MU metrics (ie, access, send, and VDT), 10 studies (13%) out of the 76 studies conducted in the United States did not meet at least one of the MU metrics, meaning that 66 studies (87%) did analyze MU criteria in some capacity, irrespective of an explicit motive. However, only 18 (24%) of the 76 US studies explicitly implicated MU criteria as a driving force behind their study.

Further data analysis revealed that a larger percentage of manuscripts investigated three (30%) or four (26%) metrics rather than two metrics (21%), nodding to the perceived

complexity of the relationship between variables influencing portal utilization. Only one study, by Neuner and colleagues [42], investigated all six variables included in MU guidelines, focusing on investigating enrollment and use based on MU guidelines, as well as satisfaction, highlighting the lack of exhaustive analysis of all available metrics. While "use" in this manuscript was defined as patients accessing their patient portal, use in other studies has been defined as number of enrollees [32], access plus protocol-specific assessments and secure messaging [43], percentage with at least one login [44], and number of patients who viewed physician notes within 30 days of their visit [45]. While Henry and colleagues [46] defined use as registration versus not, Arcury et al [47] and Graetz et al [86] analyzed use as a patient-reported binary metric, highlighting the variability in this metric's composition. The many studies stratifying use based on at least one login could be capturing the login required to create the account and not portal utilization as proxy for health care engagement [6,29,48-50]. Some of these studies created specific classifications for users, including "nonusers," "readers," and "readers and writers" [51], potentially to mitigate their definition of use. As an example of the complexity of patient portal data, in an effort to ensure accuracy of the study population, Phelps et al [52] stratified users by the absence of any portal login in the past 6 months, despite at least two lab uploads, to ensure the population studied was alive and had reason to access the portal. Others classified use by the completion of at least one survey during the study period [53], the total number of patients on the mobile app [38], the number of patients initiating the online refill function [8], or contact with the care messenger via the portal [54], highlighting the variability in the definition of this fundamental metric.

Table 4. Meaningful use (MU) definitions in US studies (n=76).

Measure	Definition	Studies with requisite measure, n (%)[a]	Study citations
Explicit MU motive	• Authors state outright that the motivation behind the study stemmed from MU criteria	18 (24)	[7,18,30,32,37,42,45,55-65]
Mention of MU	• Some mention of MU was made in the introduction or discussion	40 (53)	[6,8-10,33,36,38-40,43,44,47,50,53,54,66-90]
Access	• Patient login into the portal	45 (59)	[6,9,18,30,32,33,36-38,40,42-45,50,51,53-56,58,60-66,69,71,72,80-82,85,87-96]
Send	• Sending a secure message to the health care provider	45 (59)	[8-10,18,30,32,35,39,40,42-44,50,51,54-60,62,63,65-67,70-73,75-78,85-88,91,94-99]
View-download-transmit (VDT)	• Ability to view, download, and transmit health information within 4 days of an office visit (providers) or 36 hours of discharge (hospital)	9 (12)	[36,42,56,60,61,65,72,81,96]
MU-2	• Met requirements for access, send, and VDT	6 (8)	[42,56,60,65,72,96]

[a]Percentages exceed 100% total because studies could meet more than one criterion, and MU-2 represents studies that met all three conditions (access, send, and VDT) that together entail the full requirements of Centers for Medicare & Medicaid Services MU stage 2.

Lacking standardized definitions, variable consistency in the application of MU terminology appears throughout the published literature [2,32,43,106]. At the most fundamental level, the lack of distinction between the patient health record (PHR), whose ownership and management lies with the patient, and the patient portal, whose ownership and management lies with the health care organization, was evident in publications that investigated patient portals but included information on PHRs in the statistical analysis [13,104,107]. Further, Devkota and colleagues [51] highlighted that mixed outcomes regarding the relationship between frequency of portal utilization and health outcomes are rooted in how studies analyze portal interaction; while some studies focused on message counts [31,41,70], others focused on interaction intensity with providers through portal messaging, stratifying by no use, read-only, and read-and-write [18,43]. Further, Jones et al [71] included consistency as part of their frequency measurement, an inclusion not found in other manuscripts. Baldwin and colleagues [83] explicitly stated in their manuscript that "registration rates and ID verifications do not account for the people who register but do not actively use the portal," citing difficulties in their "use" analysis from patients who "report login issues and difficulty navigating portals."

Discussion

Principal Findings

Portal analyses have extended beyond MU-2 criteria in an effort to best meet provider and patient needs. Fast Pass—an "automated rescheduling system" requiring opt-in through a patient portal—not only indirectly measured patient use/adoption through logins to enter the program but showed that automated

rescheduling prompts reduced no-show appointments by 38% [108]. Patient portal utilization extending beyond MU criteria has been critical during the COVID-19 pandemic, with Patel et al [109] describing how pediatric patient portals pivoted from traditional, in-person enrollment methods and Judson et al [110] detailing the creation of a COVID-19 self-triage and assessment tool for primary care patients. The widespread patient portal adoption in the United States provides the necessary foundation for patients to access telemedicine visits while simultaneously creating a digital divide—a topic vastly cited since the emergence of patient portals [111-113]. Graetz et al [86] asserted that the digital divide particularly impacts disadvantaged groups; they observed that the use of a personal computer and internet access "explained 52% of the association between race and secure message use and 60% of the association between income and use," and suggested that providing portal access across multiple platforms, including telephones, could reduce message use disparities. Further, initiation of portal use has been found to be "lower for racial and ethnic minorities, persons of lower socioeconomic status, and those without neighborhood broadband internet access," leading to a digital divide in portal utilization [63]. A mismatch between "MU-based metrics of patient engagement and the priorities and needs of safety net populations" has also been cited [60], mirroring the recent combination of patient and provider use by Margolius et al [103] that found an increased quantity of patient-driven messages received by clinicians with a more robust or wealthier patient base. We recognize that portal utilization varies widely across institutions, with some institutions using patient portals for appointment scheduling, uploading demographic data and completing assessments prior to clinical visits, and even downloading parking passes for on-site visits, while other institutions emphasize portal use more heavily for message utilization and/or lab results. While our investigation was focused primarily on definitional differences in use across institutions, future investigations should explore portal-specific patient education and training, along with differences in portal functionality.

Recognizing that the intended target/purpose of portal interventions varies widely, it follows that the patient portal metrics utilized will be based on the functionalities being tested in each intervention. This fact results in the inability to generate specific overarching recommendations regarding portal analysis; however, systematic analysis of portal functions using clear definitions provides a foundation from which future studies can more readily compare portal use. Further, defining "use" more substantively by removing the baseline of single login—which could be the login used to create the portal itself without meaningful interaction with portal functions—could further facilitate the generation of meaningful utilization data. Therefore, we recommend that future studies clearly and specifically define the portal metrics utilized to allow for comparisons across studies and avoid using a binary measure of patient use/adoption that includes just one login, as the creation of a portal account often requires an initial login that does not necessarily equate to any MU. Relatedly, Gheorghiu and Hagens [106] criticized analyzing only the aggregate number of portal accesses because this cannot distinguish between a large yet infrequent number of users from a few frequent users. Further, we recommend that all future patient portal studies include the following population characteristics: the total organization population that could have access to the portal; the number of patients (and percentage of total) that currently have a patient portal account (regardless of use); and the number of patients that have used their account within the past year, with use being defined as two logins. This information will allow for meaningful comparisons across studies without being overly cumbersome to attain.

The diversity of metrics found in this review may also inform patient portal operations and dashboards of what may be worth tracking for research purposes. Additionally, the categories that emerged during this review could be used going forward to classify the variables of interest in future patient portal studies (ie, patient and/or provider use/adoption, duration, frequency, intensity, and super user). A few articles noted in the study limitations that a metric the authors deemed valuable to report could not be tracked for lack of available data. These included data on frequency for early adopters [84], intensity in the form of portal components accessed [35], and patient access to radiology images—the importance of which the authors noted was independent of MU-2 compliance, but, being unrelated to compliance, was not available [64]. One cannot study—or improve—what one does not track or offer to patients.

Further, the stratification of patient use/adoption provides an important area for future analysis. For instance, Zhong et al [79] cited the lack of quantification of active use (eg, in terms of per-user frequency or by-function intensity) as a study limitation. Some studies analyzed portal utilization through more "active" measures using a variety of multicomponent or stratified criteria. For instance, Devkota et al [51] grouped patients into "nonusers" who either did not activate their account or activated the account but did not write a message, "readers" who accessed but did not reply to emails, and "readers and writers" who read and subsequently wrote emails [51]. Oest et al [80] delineated commonly accessed portal features, including access to outpatient laboratory and radiology results. Miles et al [82] stratified use by types of available reports accessed—that is, the percentage of patients who viewed their radiology results were compared with views for other reports among all portal-registered patients. Manard et al [35] delineated active versus no active use by patients who wrote messages versus those who did not register or registered but only read messages.

Examples of methods used to stratify patient use/adoption in more "passive" terms include grouping patients by login versus no login (eg, Ronda et al [48] and Price-Haywood and Luo [6]) and defining use as patients registered for notifications versus those not registered at all (Henry et al [46]). Jones et al [71] defined an "active user" as one who had at least two portal sessions over the study period, with session defined as login-to-logout or until the 20-minute timeout occurred. Mirroring this, Petullo et al [75] distinguished account activation, defined as "active," from patients sending messages, defined as "users." Both Jones et al [71] and Petullo et al [75] created an interesting dynamic in which the publication employed the terminology of "active" without necessitating further portal engagement beyond login. Further, Masterson Creber et al [7] defined their own "Patient Activation Measure"

to gauge patient engagement with portal functions, highlighting the need to more concretely define the parameters surrounding "active" versus "passive" portal use (and to disambiguate "active" as in user activity from merely "active" as in activated/registered accounts of otherwise passive users). Future measures should attempt to create a distinction between active and passive use.

Provider portal utilization drives patient utilization, with provider messaging levels and types predicting subsequent patient communication behavior [56] and provider responses to other patients' messages driving a statistically significant increase in messages initiated by their patients [59]. A total of 30 studies in our systematic review specifically analyzed provider use, with nearly half of those studies analyzing provider-initiated messages. This fact further highlights the notion of physician use predicting patient use of portal functions and the intersection of physician portal use with institutional support of portal utilization. Recognizing that organizational policies mandate physician portal use, and nonuse, future studies should examine metrics for provider use more distinctly from patient use. As mentioned previously, provider use was only analyzed separately from patient use in 6% of studies, generating an untapped lens through which to investigate the driving forces of patient portal utilization. Mafi et al [45] found that individualized physician reminders to patients alerting them of completed visit notes drove patient portal utilization and engagement. However, provider and patient utilization are also intrinsically linked: Laccetti et al's [41] analysis on portal use by clinical staff hinged on the clinical actions performed based on received patient messages, and Crotty et al's [100] analysis of characteristics of unread messages necessitated message sending by the provider and lack of message reading by the patient. A recent study by Huerta et al [105] provided guidance on patient portal log file analysis and developed a taxonomy of computed analytic metrics. Patients who utilize their portal for longer periods are more likely to prefer communication through said portal, highlighting the importance of analyzing both patient and provider utilization [114].

Conclusion

Our investigation supports the claim that not all health care systems study patient portal utilization systematically; thus, health care system support of different communication modalities is essential. Currently, the published literature is limited to analysis that is mostly based on patient portal utilization, as defined by MU criteria. More in-depth studies, mirroring the log file analysis conducted by Huerta et al [105] and, more recently, Di Tosto et al [115] that included a blueprint of individual patients' portal actions, would fulfill our endeavor to utilize patient portal data more completely than the literature currently reports routinely. A systematic approach to measurement of portal usage is necessary to more readily draw comparisons across existing and future studies. Investigation of both provider and patient use/adoption will provide insight to generate a platform that is most beneficial for all users. One important limitation to note is that our review was limited to one database, but the main outlets for patient portal studies were included. Further, this is the largest known review examining patient portal research and the only review focusing on associated MU compliance assessment. Future investigation should more holistically analyze patient portal components in combination with the utilization of health services to elicit potential relationships currently unseen between portal use and patient health outcomes and to explore use that is, in the given context, truly meaningful.

Authors' Contributions

TM conceptualized the research aim and supervised the project; TM and JMK designed the search strategy; LLB, JMK, and TM conducted literature searches and performed the data charting analysis; LLB and TM drafted the manuscript; TM, LLB, SLJ, and AK interpreted the study results; and all authors reviewed, revised, and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed summary of included studies.
[XLSX File (Microsoft Excel File), 41 KB - jmir_v23i2e23493_app1.xlsx]

References

1. Turner K, Hong Y, Yadav S, Huo J, Mainous A. Patient portal utilization: Before and after stage 2 electronic health record meaningful use. J Am Med Informatics Assoc 2019;26(10):960-967. [doi: 10.1093/jamia/ocz030]
2. Irizarry T, DeVito DA, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015 Jun;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
3. Lockwood MB, Dunn-Lopez K, Pauls H, Burke L, Shah SD, Saunders MA. If you build it, they may not come: modifiable barriers to patient portal use among pre- and post-kidney transplant patients. JAMIA Open 2018 Oct;1(2):255-264 [FREE Full text] [doi: 10.1093/jamiaopen/ooy024] [Medline: 31984337]
4. Weitzman ER, Kaci L, Mandl KD. Acceptability of a personally controlled health record in a community-based setting: implications for policy and design. J Med Internet Res 2009 Apr 29;11(2):e14 [FREE Full text] [doi: 10.2196/jmir.1187] [Medline: 19403467]

5. Heath S. Patient Portal Adoption Tops 90%, But Strong Patient Use Is Needed. Patient Engagement HIT. URL: https://patientengagementhit.com/news/patient-portal-adoption-tops-90-but-strong-patient-use-is-needed [accessed 2020-04-29]
6. Price-Haywood EG, Luo Q. Primary Care Practice Reengineering and Associations With Patient Portal Use, Service Utilization, and Disease Control Among Patients With Hypertension and/or Diabetes. Ochsner J 2017;17(1):103-111 [FREE Full text] [Medline: 28331456]
7. Masterson Creber R, Prey J, Ryan B, Alarcon I, Qian M, Bakken S, et al. Engaging hospitalized patients in clinical care: Study protocol for a pragmatic randomized controlled trial. Contemp Clin Trials 2016 Mar;47:165-171 [FREE Full text] [doi: 10.1016/j.cct.2016.01.005] [Medline: 26795675]
8. Sarkar U, Lyles CR, Parker MM, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. Med Care 2014 Mar;52(3):194-201 [FREE Full text] [doi: 10.1097/MLR.0000000000000069] [Medline: 24374412]
9. Lyles CR, Sarkar U, Schillinger D, Ralston JD, Allen JY, Nguyen R, et al. Refilling medications through an online patient portal: consistent improvements in adherence across racial/ethnic groups. J Am Med Inform Assoc 2016 Apr;23(e1):e28-e33 [FREE Full text] [doi: 10.1093/jamia/ocv126] [Medline: 26335983]
10. Tulu B, Trudel J, Strong DM, Johnson SA, Sundaresan D, Garber L. Patient Portals: An Underused Resource for Improving Patient Engagement. Chest 2016 Jan;149(1):272-277. [doi: 10.1378/chest.14-2559] [Medline: 26066707]
11. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient Portals Facilitating Engagement With Inpatient Electronic Medical Records: A Systematic Review. J Med Internet Res 2019 Apr 11;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
12. McAlearney AS, Sieck CJ, Hefner JL, Aldrich AM, Walker DM, Rizer MK, et al. High Touch and High Tech (HT2) Proposal: Transforming Patient Engagement Throughout the Continuum of Care by Engaging Patients with Portal Technology at the Bedside. JMIR Res Protoc 2016 Nov 29;5(4):e221 [FREE Full text] [doi: 10.2196/resprot.6355] [Medline: 27899338]
13. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
14. Sulieman L, Yin Z, Malin BA. Why Patient Portal Messages Indicate Risk of Readmission for Patients with Ischemic Heart Disease. AMIA Annu Symp Proc 2019;2019:828-837 [FREE Full text] [Medline: 32308879]
15. Irizarry T, Shoemake J, Nilsen ML, Czaja S, Beach S, DeVito Dabbs A. Patient Portals as a Tool for Health Care Engagement: A Mixed-Method Study of Older Adults With Varying Levels of Health Literacy and Prior Patient Portal Use. J Med Internet Res 2017 Mar 30;19(3):e99 [FREE Full text] [doi: 10.2196/jmir.7099] [Medline: 28360022]
16. Wright A, Feblowitz J, Samal L, McCoy AB, Sittig DF. The Medicare Electronic Health Record Incentive Program: provider performance on core and menu measures. Health Serv Res 2014 Feb;49(1 Pt 2):325-346 [FREE Full text] [doi: 10.1111/1475-6773.12134] [Medline: 24359554]
17. Levine DM, Healey MJ, Wright A, Bates DW, Linder JA, Samal L. Changes in the quality of care during progress from stage 1 to stage 2 of Meaningful Use. J Am Med Inform Assoc 2017 Mar 01;24(2):394-397 [FREE Full text] [doi: 10.1093/jamia/ocw127] [Medline: 27567000]
18. Griffin A, Skinner A, Thornhill J, Weinberger M. Patient Portals: Who uses them? What features do they use? And do they reduce hospital readmissions? Appl Clin Inform 2016;7(2):489-501 [FREE Full text] [doi: 10.4338/ACI-2016-01-RA-0003] [Medline: 27437056]
19. Ford EW, Hesse BW, Huerta TR. Personal Health Record Use in the United States: Forecasting Future Adoption Levels. J Med Internet Res 2016 Mar 30;18(3):e73 [FREE Full text] [doi: 10.2196/jmir.4973] [Medline: 27030105]
20. Powell KR. Patient-Perceived Facilitators of and Barriers to Electronic Portal Use: A Systematic Review. Comput Inform Nurs 2017 Nov;35(11):565-573. [doi: 10.1097/CIN.0000000000000377] [Medline: 28723832]
21. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
22. Megan Forster M, Dennison K, Callen J, Andrew A, Westbrook JI. Maternity patients' access to their electronic medical records: use and perspectives of a patient portal. Health Inf Manag 2015;44(1):4-11. [doi: 10.1177/183335831504400101] [Medline: 27092464]
23. Garrido T, Kanter M, Meng D, Turley M, Wang J, Sue V, et al. Race/ethnicity, personal health record access, and quality of care. Am J Manag Care 2015 Feb 01;21(2):e103-e113 [FREE Full text] [Medline: 25880485]
24. Gheorghiu B, Hagens S. Use and maturity of electronic patient portals. Stud Health Technol Inform 2017;234:136-141. [doi: 10.3233/978-1-61499-742-9-136] [Medline: 28186030]
25. Kipping S, Stuckey MI, Hernandez A, Nguyen T, Riahi S. A Web-Based Patient Portal for Mental Health Care: Benefits Evaluation. J Med Internet Res 2016 Nov 16;18(11):e294 [FREE Full text] [doi: 10.2196/jmir.6483] [Medline: 27852556]
26. Redelmeier DA, Kraus NC. Patterns in Patient Access and Utilization of Online Medical Records: Analysis of MyChart. J Med Internet Res 2018 Feb 06;20(2):e43 [FREE Full text] [doi: 10.2196/jmir.8372] [Medline: 29410386]
27. Riippa I, Linna M, Rönkkö I, Kröger V. Use of an electronic patient portal among the chronically ill: an observational study. J Med Internet Res 2014 Dec 08;16(12):e275 [FREE Full text] [doi: 10.2196/jmir.3722] [Medline: 25488754]

28. Runaas L, Hanauer D, Maher M, Bischoff E, Fauer A, Hoang T, et al. BMT Roadmap: A User-Centered Design Health Information Technology Tool to Promote Patient-Centered Care in Pediatric Hematopoietic Cell Transplantation. Biol Blood Marrow Transplant 2017 May;23(5):813-819 [FREE Full text] [doi: 10.1016/j.bbmt.2017.01.080] [Medline: 28132870]
29. Lau M, Campbell H, Tang T, Thompson DJS, Elliott T. Impact of patient use of an online patient portal on diabetes outcomes. Can J Diabetes 2014 Feb;38(1):17-21. [doi: 10.1016/j.jcjd.2013.10.005] [Medline: 24485208]
30. Emani S, Healey M, Ting DY, Lipsitz SR, Ramelson H, Suric V, et al. Awareness and Use of the After-Visit Summary Through a Patient Portal: Evaluation of Patient Characteristics and an Application of the Theory of Planned Behavior. J Med Internet Res 2016 Apr 13;18(4):e77 [FREE Full text] [doi: 10.2196/jmir.5207] [Medline: 27076485]
31. Boogerd E, Maas-Van Schaaijk NM, Sas TC, Clement-de Boers A, Smallenbroek M, Nuboer R, et al. Sugarsquare, a Web-Based Patient Portal for Parents of a Child With Type 1 Diabetes: Multicenter Randomized Controlled Feasibility Trial. J Med Internet Res 2017 Aug 22;19(8):e287 [FREE Full text] [doi: 10.2196/jmir.6639] [Medline: 28830853]
32. Thompson LA, Martinko T, Budd P, Mercado R, Schentrup AM. Meaningful Use of a Confidential Adolescent Patient Portal. J Adolesc Health 2016 Feb;58(2):134-140. [doi: 10.1016/j.jadohealth.2015.10.015] [Medline: 26802988]
33. Woods SS, Forsberg CW, Schwartz EC, Nazi KM, Hibbard JH, Houston TK, et al. The Association of Patient Factors, Digital Access, and Online Behavior on Sustained Patient Portal Use: A Prospective Cohort of Enrolled Users. J Med Internet Res 2017 Oct 17;19(10):e345 [FREE Full text] [doi: 10.2196/jmir.7895] [Medline: 29042345]
34. King G, Maxwell J, Karmali A, Hagens S, Pinto M, Williams L, et al. Connecting Families to Their Health Record and Care Team: The Use, Utility, and Impact of a Client/Family Health Portal at a Children's Rehabilitation Hospital. J Med Internet Res 2017 Apr 06;19(4):e97 [FREE Full text] [doi: 10.2196/jmir.6811] [Medline: 28385680]
35. Manard W, Scherrer JF, Salas J, Schneider FD. Patient Portal Use and Blood Pressure Control in Newly Diagnosed Hypertension. J Am Board Fam Med 2016;29(4):452-459 [FREE Full text] [doi: 10.3122/jabfm.2016.04.160008] [Medline: 27390376]
36. Krasowski MD, Grieme CV, Cassady B, Dreyer NR, Wanat KA, Hightower M, et al. Variation in Results Release and Patient Portal Access to Diagnostic Test Results at an Academic Medical Center. J Pathol Inform 2017;8:45 [FREE Full text] [doi: 10.4103/jpi.jpi_53_17] [Medline: 29226008]
37. Pillemer F, Price RA, Paone S, Martich GD, Albert S, Haidari L, et al. Direct Release of Test Results to Patients Increases Patient Engagement and Utilization of Care. PLoS One 2016;11(6):e0154743 [FREE Full text] [doi: 10.1371/journal.pone.0154743] [Medline: 27337092]
38. Pearl R. Kaiser Permanente Northern California: current experiences with internet, mobile, and video technologies. Health Aff (Millwood) 2014 Feb;33(2):251-257. [doi: 10.1377/hlthaff.2013.1005] [Medline: 24493768]
39. Garrido T, Meng D, Wang JJ, Palen TE, Kanter MH. Secure e-mailing between physicians and patients: transformational change in ambulatory care. J Ambul Care Manage 2014;37(3):211-218 [FREE Full text] [doi: 10.1097/JAC.0000000000000043] [Medline: 24887522]
40. Toscos T, Daley C, Heral L, Doshi R, Chen Y, Eckert GJ, et al. Impact of electronic personal health record use on engagement and intermediate health outcomes among cardiac patients: a quasi-experimental study. J Am Med Inform Assoc 2016 Jan;23(1):119-128 [FREE Full text] [doi: 10.1093/jamia/ocv164] [Medline: 26912538]
41. Laccetti AL, Chen B, Cai J, Gates S, Xie Y, Lee SJC, et al. Increase in Cancer Center Staff Effort Related to Electronic Patient Portal Use. J Oncol Pract 2016 Dec;12(12):e981-e990 [FREE Full text] [doi: 10.1200/JOP.2016.011817] [Medline: 27601511]
42. Neuner J, Fedders M, Caravella M, Bradford L, Schapira M. Meaningful use and the patient portal: patient enrollment, use, and satisfaction with patient portals at a later-adopting center. Am J Med Qual 2015;30(2):105-113 [FREE Full text] [doi: 10.1177/1062860614523488] [Medline: 24563085]
43. Sorondo B, Allen A, Fathima S, Bayleran J, Sabbagh I. Patient Portal as a Tool for Enhancing Patient Experience and Improving Quality of Care in Primary Care Practices. EGEMS (Wash DC) 2016;4(1):1262 [FREE Full text] [doi: 10.13063/2327-9214.1262] [Medline: 28203611]
44. Jhamb M, Cavanaugh KL, Bian A, Chen G, Ikizler TA, Unruh ML, et al. Disparities in Electronic Health Record Patient Portal Use in Nephrology Clinics. Clin J Am Soc Nephrol 2015 Nov 06;10(11):2013-2022 [FREE Full text] [doi: 10.2215/CJN.01640215] [Medline: 26493242]
45. Mafi JN, Mejilla R, Feldman H, Ngo L, Delbanco T, Darer J, et al. Patients learning to read their doctors' notes: the importance of reminders. J Am Med Inform Assoc 2016 Sep;23(5):951-955 [FREE Full text] [doi: 10.1093/jamia/ocv167] [Medline: 26911830]
46. Henry SL, Shen E, Ahuja A, Gould MK, Kanter MH. The Online Personal Action Plan: A Tool to Transform Patient-Enabled Preventive and Chronic Care. Am J Prev Med 2016 Jul;51(1):71-77. [doi: 10.1016/j.amepre.2015.11.014] [Medline: 26826751]
47. Arcury TA, Quandt SA, Sandberg JC, Miller DP, Latulipe C, Leng X, et al. Patient Portal Utilization Among Ethnically Diverse Low Income Older Adults: Observational Study. JMIR Med Inform 2017 Nov 14;5(4):e47 [FREE Full text] [doi: 10.2196/medinform.8026] [Medline: 29138129]

48. Ronda MCM, Dijkhorst-Oei L, Rutten GEHM. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014 Nov 25;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
49. Groen WG, Kuijpers W, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Supporting Lung Cancer Patients With an Interactive Patient Portal: Feasibility Study. JMIR Cancer 2017 Aug 08;3(2):e10 [FREE Full text] [doi: 10.2196/cancer.7443] [Medline: 28790025]
50. Bell K, Warnick E, Nicholson K, Ulcoq S, Kim SJ, Schroeder GD, et al. Patient Adoption and Utilization of a Web-Based and Mobile-Based Portal for Collecting Outcomes After Elective Orthopedic Surgery. Am J Med Qual 2018;33(6):649-656. [doi: 10.1177/1062860618765083] [Medline: 29562769]
51. Devkota B, Salas J, Sayavong S, Scherrer JF. Use of an Online Patient Portal and Glucose Control in Primary Care Patients with Diabetes. Popul Health Manag 2016 Apr;19(2):125-131. [doi: 10.1089/pop.2015.0034] [Medline: 26237200]
52. Phelps RG, Taylor J, Simpson K, Samuel J, Turner AN. Patients' continuing use of an online health record: a quantitative evaluation of 14,000 patient years of access data. J Med Internet Res 2014 Oct 24;16(10):e241 [FREE Full text] [doi: 10.2196/jmir.3371] [Medline: 25344621]
53. Fiks AG, DuRivage N, Mayne SL, Finch S, Ross ME, Giacomini K, et al. Adoption of a Portal for the Primary Care Management of Pediatric Asthma: A Mixed-Methods Implementation Study. J Med Internet Res 2016 Jun 29;18(6):e172 [FREE Full text] [doi: 10.2196/jmir.5610] [Medline: 27357835]
54. Pecina J, North F, Williams MD, Angstman KB. Use of an on-line patient portal in a depression collaborative care management program. J Affect Disord 2017 Jan 15;208:1-5. [doi: 10.1016/j.jad.2016.08.034] [Medline: 27736688]
55. Williamson RS, Cherven BO, Gilleland Marchak J, Edwards P, Palgon M, Escoffery C, et al. Meaningful Use of an Electronic Personal Health Record (ePHR) among Pediatric Cancer Survivors. Appl Clin Inform 2017 Mar 15;8(1):250-264 [FREE Full text] [doi: 10.4338/ACI-2016-11-RA-0189] [Medline: 28293684]
56. Kamo N, Bender AJ, Kalmady K, Blackmore CC. Meaningful use of the electronic patient portal - Virginia Mason's journey to create the perfect online patient experience. Healthc (Amst) 2017 Dec;5(4):221-226. [doi: 10.1016/j.hjdsi.2016.09.003] [Medline: 27727028]
57. Reed M, Graetz I, Gordon N, Fung V. Patient-initiated e-mails to providers: associations with out-of-pocket visit costs, and impact on care-seeking and health. Am J Manag Care 2015 Dec 01;21(12):e632-e639 [FREE Full text] [Medline: 26760425]
58. Crotty BH, Tamrat Y, Mostaghimi A, Safran C, Landon BE. Patient-to-physician messaging: volume nearly tripled as more patients joined system, but per capita rate plateaued. Health Aff (Millwood) 2014 Oct;33(10):1817-1822 [FREE Full text] [doi: 10.1377/hlthaff.2013.1145] [Medline: 25288428]
59. Wolcott V, Agarwal R, Nelson DA. Is Provider Secure Messaging Associated With Patient Messaging Behavior? Evidence From the US Army. J Med Internet Res 2017 Apr 06;19(4):e103 [FREE Full text] [doi: 10.2196/jmir.6804] [Medline: 28385681]
60. Ackerman SL, Sarkar U, Tieu L, Handley MA, Schillinger D, Hahn K, et al. Meaningful use in the safety net: a rapid ethnography of patient portal implementation at five community health centers in California. J Am Med Inform Assoc 2017 Sep 01;24(5):903-912 [FREE Full text] [doi: 10.1093/jamia/ocx015] [Medline: 28340229]
61. Krist AH, Woolf SH, Bello GA, Sabo RT, Longo DR, Kashiri P, et al. Engaging primary care patients to use a patient-centered personal health record. Ann Fam Med 2014;12(5):418-426 [FREE Full text] [doi: 10.1370/afm.1691] [Medline: 25354405]
62. Gordon NP, Hornbrook MC. Differences in Access to and Preferences for Using Patient Portals and Other eHealth Technologies Based on Race, Ethnicity, and Age: A Database and Survey Study of Seniors in a Large Health Plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
63. Perzynski AT, Roach MJ, Shick S, Callahan B, Gunzler D, Cebul R, et al. Patient portals and broadband internet inequality. J Am Med Inform Assoc 2017 Sep 01;24(5):927-932 [FREE Full text] [doi: 10.1093/jamia/ocx020] [Medline: 28371853]
64. Reicher JJ, Reicher MA. Implementation of Certified EHR, Patient Portal, and "Direct" Messaging Technology in a Radiology Environment Enhances Communication of Radiology Results to Both Referring Physicians and Patients. J Digit Imaging 2016 Jun;29(3):337-340 [FREE Full text] [doi: 10.1007/s10278-015-9845-x] [Medline: 26588906]
65. Shaw CL, Casterline GL, Taylor D, Fogle M, Granger B. Increasing Health Portal Utilization in Cardiac Ambulatory Patients: A Pilot Project. Comput Inform Nurs 2017 Oct;35(10):512-519. [doi: 10.1097/CIN.0000000000000361] [Medline: 28541965]
66. Weisner CM, Chi FW, Lu Y, Ross TB, Wood SB, Hinman A, et al. Examination of the Effects of an Intervention Aiming to Link Patients Receiving Addiction Treatment With Health Care: The LINKAGE Clinical Trial. JAMA Psychiatry 2016 Aug 01;73(8):804-814 [FREE Full text] [doi: 10.1001/jamapsychiatry.2016.0970] [Medline: 27332703]
67. Bose-Brill S, Feeney M, Prater L, Miles L, Corbett A, Koesters S. Validation of a Novel Electronic Health Record Patient Portal Advance Care Planning Delivery System. J Med Internet Res 2018 Jun 26;20(6):e208 [FREE Full text] [doi: 10.2196/jmir.9203] [Medline: 29945860]
68. Fiks AG, Mayne SL, Karavite DJ, Suh A, O'Hara R, Localio AR, et al. Parent-reported outcomes of a shared decision-making portal in asthma: a practice-based RCT. Pediatrics 2015 Apr;135(4):e965-e973 [FREE Full text] [doi: 10.1542/peds.2014-3167] [Medline: 25755233]

69. Aljabri D, Dumitrascu A, Burton MC, White L, Khan M, Xirasagar S, et al. Patient portal adoption and use by hospitalized cancer patients: a retrospective study of its impact on adverse events, utilization, and patient satisfaction. BMC Med Inform Decis Mak 2018 Jul 27;18(1):70 [FREE Full text] [doi: 10.1186/s12911-018-0644-4] [Medline: 30053809]
70. Chung S, Panattoni L, Chi J, Palaniappan L. Can Secure Patient-Provider Messaging Improve Diabetes Care? Diabetes Care 2017 Oct;40(10):1342-1348. [doi: 10.2337/dc17-0140] [Medline: 28807977]
71. Jones JB, Weiner JP, Shah NR, Stewart WF. The wired patient: patterns of electronic patient portal use among patients with cardiac disease or diabetes. J Med Internet Res 2015 Feb 20;17(2):e42 [FREE Full text] [doi: 10.2196/jmir.3157] [Medline: 25707036]
72. Kelly MM, Hoonakker PLT, Dean SM. Using an inpatient portal to engage families in pediatric hospital care. J Am Med Inform Assoc 2017 Jan;24(1):153-161 [FREE Full text] [doi: 10.1093/jamia/ocw070] [Medline: 27301746]
73. Haun JN, Lind JD, Shimada SL, Martin TL, Gosline RM, Antinori N, et al. Evaluating user experiences of the secure messaging tool on the Veterans Affairs' patient portal system. J Med Internet Res 2014 Mar 06;16(3):e75 [FREE Full text] [doi: 10.2196/jmir.2976] [Medline: 24610454]
74. Mook PJ, Trickey AW, Krakowski KE, Majors S, Theiss MA, Fant C, et al. Exploration of Portal Activation by Patients in a Healthcare System. Comput Inform Nurs 2018 Jan;36(1):18-26. [doi: 10.1097/CIN.0000000000000392] [Medline: 29049084]
75. Petullo B, Noble B, Dungan KM. Effect of Electronic Messaging on Glucose Control and Hospital Admissions Among Patients with Diabetes. Diabetes Technol Ther 2016 Sep;18(9):555-560 [FREE Full text] [doi: 10.1089/dia.2016.0105] [Medline: 27398824]
76. Shenson JA, Cronin RM, Davis SE, Chen Q, Jackson GP. Rapid growth in surgeons' use of secure messaging in a patient portal. Surg Endosc 2016 Apr;30(4):1432-1440 [FREE Full text] [doi: 10.1007/s00464-015-4347-y] [Medline: 26123340]
77. Dexter EN, Fields S, Rdesinski RE, Sachdeva B, Yamashita D, Marino M. Patient-Provider Communication: Does Electronic Messaging Reduce Incoming Telephone Calls? J Am Board Fam Med 2016;29(5):613-619 [FREE Full text] [doi: 10.3122/jabfm.2016.05.150371] [Medline: 27613794]
78. Mikles SP, Mielenz TJ. Characteristics of electronic patient-provider messaging system utilisation in an urban health care organisation. J Innov Health Inform 2014 Dec 18;22(1):214-221 [FREE Full text] [doi: 10.14236/jhi.v22i1.75] [Medline: 25924551]
79. Zhong X, Liang M, Sanchez R, Yu M, Budd PR, Sprague JL, et al. On the effect of electronic patient portal on primary care utilization and appointment adherence. BMC Med Inform Decis Mak 2018 Oct 16;18(1):84 [FREE Full text] [doi: 10.1186/s12911-018-0669-8] [Medline: 30326876]
80. Oest S, Hightower M, Krasowski M. Activation and Utilization of an Electronic Health Record Patient Portal at an Academic Medical Center-Impact of Patient Demographics and Geographic Location. Acad Pathol 2018;5:2374289518797573 [FREE Full text] [doi: 10.1177/2374289518797573] [Medline: 30302394]
81. Robinson JR, Davis SE, Cronin RM, Jackson GP. Use of a Patient Portal During Hospital Admissions to Surgical Services. AMIA Annu Symp Proc 2016;2016:1967-1976 [FREE Full text] [Medline: 28269956]
82. Miles RC, Hippe DS, Elmore JG, Wang CL, Payne TH, Lee CI. Patient Access to Online Radiology Reports: Frequency and Sociodemographic Characteristics Associated with Use. Acad Radiol 2016 Sep;23(9):1162-1169. [doi: 10.1016/j.acra.2016.05.005] [Medline: 27287715]
83. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: Pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2017 Sep;5(3):81-85 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]
84. Payne TH, Beahan S, Fellner J, Martin D, Elmore JG. Health Records All Access Pass. Patient Portals That Allow Viewing of Clinical Notes and Hospital Discharge Summaries: The University of Washington Opennotes Implementation Experience. J AHIMA 2016 Aug;87(8):36-39. [Medline: 29425012]
85. Smith SG, O'Conor R, Aitken W, Curtis LM, Wolf MS, Goel MS. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015 Jul;22(4):888-895 [FREE Full text] [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
86. Graetz I, Gordon N, Fung V, Hamity C, Reed ME. The Digital Divide and Patient Portals: Internet Access Explained Differences in Patient Portal Use for Secure Messaging by Age, Race, and Income. Med Care 2016 Aug;54(8):772-779. [doi: 10.1097/MLR.0000000000000560] [Medline: 27314262]
87. Steitz B, Cronin R, Davis S, Yan E, Jackson G. Long-term Patterns of Patient Portal Use for Pediatric Patients at an Academic Medical Center. Appl Clin Inform 2017 Aug 02;8(3):779-793 [FREE Full text] [doi: 10.4338/ACI-2017-01-RA-0005] [Medline: 28765865]
88. Wallace LS, Angier H, Huguet N, Gaudino JA, Krist A, Dearing M, et al. Patterns of Electronic Portal Use among Vulnerable Patients in a Nationwide Practice-based Research Network: From the OCHIN Practice-based Research Network (PBRN). J Am Board Fam Med 2016;29(5):592-603 [FREE Full text] [doi: 10.3122/jabfm.2016.05.160046] [Medline: 27613792]
89. Bajracharya AS, Crotty BH, Kowaloff HB, Safran C, Slack WV. Improving health care proxy documentation using a web-based interview through a patient portal. J Am Med Inform Assoc 2016 May;23(3):580-587 [FREE Full text] [doi: 10.1093/jamia/ocv133] [Medline: 26568608]

90. Vydra TP, Cuaresma E, Kretovics M, Bose-Brill S. Diffusion and Use of Tethered Personal Health Records in Primary Care. Perspect Health Inf Manag 2015;12:1c [FREE Full text] [Medline: 26755897]
91. Peremislov D. Patient Use of the Electronic Communication Portal in Management of Type 2 Diabetes. Comput Inform Nurs 2017 Sep;35(9):473-482. [doi: 10.1097/CIN.0000000000000348] [Medline: 28323648]
92. Ahmedani BK, Belville-Robertson T, Hirsch A, Jurayj A. An Online Mental Health and Wellness Intervention Supplementing Standard Care of Depression and Anxiety. Arch Psychiatr Nurs 2016 Dec;30(6):666-670. [doi: 10.1016/j.apnu.2016.03.003] [Medline: 27888957]
93. Bower JK, Bollinger CE, Foraker RE, Hood DB, Shoben AB, Lai AM. Active Use of Electronic Health Records (EHRs) and Personal Health Records (PHRs) for Epidemiologic Research: Sample Representativeness and Nonresponse Bias in a Study of Women During Pregnancy. EGEMS (Wash DC) 2017;5(1):1263 [FREE Full text] [doi: 10.13063/2327-9214.1263] [Medline: 28303255]
94. Wolff JL, Berger A, Clarke D, Green JA, Stametz R, Yule C, et al. Patients, care partners, and shared access to the patient portal: online practices at an integrated health system. J Am Med Inform Assoc 2016 Nov;23(6):1150-1158. [doi: 10.1093/jamia/ocw025] [Medline: 27026614]
95. Shimada SL, Allison JJ, Rosen AK, Feng H, Houston TK. Sustained Use of Patient Portal Features and Improvements in Diabetes Physiological Measures. J Med Internet Res 2016 Jul 01;18(7):e179 [FREE Full text] [doi: 10.2196/jmir.5663] [Medline: 27369696]
96. Davis SE, Osborn CY, Kripalani S, Goggins KM, Jackson GP. Health Literacy, Education Levels, and Patient Portal Usage During Hospitalizations. AMIA Annu Symp Proc 2015;2015:1871-1880 [FREE Full text] [Medline: 26958286]
97. Quinn CC, Butler EC, Swasey KK, Shardell MD, Terrin MD, Barr EA, et al. Mobile Diabetes Intervention Study of Patient Engagement and Impact on Blood Glucose: Mixed Methods Analysis. JMIR Mhealth Uhealth 2018 Feb 02;6(2):e31 [FREE Full text] [doi: 10.2196/mhealth.9265] [Medline: 29396389]
98. Alpert JM, Dyer KE, Lafata JE. Patient-centered communication in digital medical encounters. Patient Educ Couns 2017 Oct;100(10):1852-1858 [FREE Full text] [doi: 10.1016/j.pec.2017.04.019] [Medline: 28522229]
99. Masterman M, Cronin RM, Davis SE, Shenson JA, Jackson GP. Adoption of Secure Messaging in a Patient Portal across Pediatric Specialties. AMIA Annu Symp Proc 2016;2016:1930-1939 [FREE Full text] [Medline: 28269952]
100. Crotty B, Mostaghimi A, O'Brien J, Bajracharya A, Safran C, Landon B. Prevalence and Risk Profile Of Unread Messages To Patients In A Patient Web Portal. Appl Clin Inform 2015;6(2):375-382 [FREE Full text] [doi: 10.4338/ACI-2015-01-CR-0006] [Medline: 26171082]
101. Dalal AK, Dykes PC, Collins S, Lehmann LS, Ohashi K, Rozenblum R, et al. A web-based, patient-centered toolkit to engage patients and caregivers in the acute care setting: a preliminary evaluation. J Am Med Inform Assoc 2016 Jan;23(1):80-87 [FREE Full text] [doi: 10.1093/jamia/ocv093] [Medline: 26239859]
102. North F, Crane SJ, Chaudhry R, Ebbert JO, Ytterberg K, Tulledge-Scheitel SM, et al. Impact of patient portal secure messages and electronic visits on adult primary care office visits. Telemed J E Health 2014 Mar;20(3):192-198 [FREE Full text] [doi: 10.1089/tmj.2013.0097] [Medline: 24350803]
103. Margolius D, Siff J, Teng K, Einstadter D, Gunzler D, Bolen S. Primary Care Physician Factors Associated with Inbox Message Volume. J Am Board Fam Med 2020;33(3):460-462 [FREE Full text] [doi: 10.3122/jabfm.2020.03.190360] [Medline: 32430380]
104. Hulter P, Pluut B, Leenen-Brinkhuis C, de Mul M, Ahaus K, Weggelaar-Jansen AM. Adopting Patient Portals in Hospitals: Qualitative Study. J Med Internet Res 2020 May 19;22(5):e16921 [FREE Full text] [doi: 10.2196/16921] [Medline: 32427110]
105. Huerta T, Fareed N, Hefner JL, Sieck CJ, Swoboda C, Taylor R, et al. Patient Engagement as Measured by Inpatient Portal Use: Methodology for Log File Analysis. J Med Internet Res 2019 Mar 25;21(3):e10957 [FREE Full text] [doi: 10.2196/10957] [Medline: 30907733]
106. Gheorghiu B, Hagens S. Measuring interoperable EHR adoption and maturity: a Canadian example. BMC Med Inform Decis Mak 2016 Jan 25;16:8 [FREE Full text] [doi: 10.1186/s12911-016-0247-x] [Medline: 26810606]
107. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
108. Chung S, Martinez MC, Frosch DL, Jones VG, Chan AS. Patient-Centric Scheduling With the Implementation of Health Information Technology to Improve the Patient Experience and Access to Care: Retrospective Case-Control Analysis. J Med Internet Res 2020 Jun 10;22(6):e16451 [FREE Full text] [doi: 10.2196/16451] [Medline: 32519970]
109. Patel PD, Cobb J, Wright D, Turer R, Jordan T, Humphrey A, et al. Rapid development of telehealth capabilities within pediatric patient portal infrastructure for COVID-19 care: barriers, solutions, results. J Am Med Inform Assoc 2020 Jul 01;27(7):1116-1120 [FREE Full text] [doi: 10.1093/jamia/ocaa065] [Medline: 32302395]
110. Judson T, Odisho A, Neinstein A, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]

111. Zhao J, Song B, Anand E, Schwartz D, Panesar M, Jackson GP, et al. Barriers, Facilitators, and Solutions to Optimal Patient Portal and Personal Health Record Use: A Systematic Review of the Literature. AMIA Annu Symp Proc 2017;2017:1913-1922 [FREE Full text] [Medline: 29854263]
112. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 01;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
113. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15 Suppl 2:183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
114. Chavez A, Bracamonte J, Kresin M, Yardley M, Grover M. High Volume Portal Usage Impacts Practice Resources. J Am Board Fam Med 2020;33(3):452-455 [FREE Full text] [doi: 10.3122/jabfm.2020.03.190401] [Medline: 32430378]
115. Di Tosto G, McAlearney AS, Fareed N, Huerta TR. Metrics for Outpatient Portal Use Based on Log File Analysis: Algorithm Development. J Med Internet Res 2020 Jun 12;22(6):e16849 [FREE Full text] [doi: 10.2196/16849] [Medline: 32530435]

Abbreviations

CMS: Centers for Medicare & Medicaid Services
MeSH: Medical Subject Heading
MU: meaningful use
MU-2: Centers for Medicare & Medicaid Services meaningful use stage 2
PHR: patient health record
VDT: view-download-transmit

Edited by R Kukafka; submitted 13.08.20; peer-reviewed by C Lyles, M Krasowski; comments to author 08.09.20; revised version received 16.11.20; accepted 31.01.21; published 25.02.21.

Please cite as:
Beal LL, Kolman JM, Jones SL, Khleif A, Menser T
Quantifying Patient Portal Use: Systematic Review of Utilization Metrics
J Med Internet Res 2021;23(2):e23493
URL: https://www.jmir.org/2021/2/e23493
doi:10.2196/23493
PMID:33629962

Original Paper

A Virtual Reality Exergame to Engage Adolescents in Physical Activity: Mixed Methods Study Describing the Formative Intervention Development Process

Nuša Farič[1], BSc, MSc; Lee Smith[2], BSc, MSc, PhD; Adrian Hon[3], BA, MA; Henry W W Potts[4*], BA, MSc, PhD; Katie Newby[5*], BSc, MSc, PhD; Andrew Steptoe[1], BA, MA, DPhil, DSC; Abi Fisher[1*], BSc, PhD

[1]Department of Behavioural Science and Health, University College London, London, United Kingdom
[2]Faculty of Science and Engineering, Anglia Ruskin University, Cambridge, United Kingdom
[3]Six to Start, London, United Kingdom
[4]Institute of Health Informatics, University College London, London, United Kingdom
[5]Department of Psychology and Sports Science, University of Hertfordshire, Hatfield, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Nuša Farič, BSc, MSc
Department of Behavioural Science and Health
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 207 7679 4466 ext 419
Email: nusa.faric.11@ucl.ac.uk

Abstract

Background: Early adolescence (13-17 years) is a critical developmental stage for physical activity promotion. Virtual reality (VR) exergaming is a promising intervention strategy to engage adolescents in physical activity.

Objective: The vEngage project aims to develop a physical activity intervention for adolescents using VR exergaming. Here, we describe the formative intervention development work and process of academic-industry collaboration.

Methods: The formative development was guided by the Medical Research Council framework and included recruiting an adolescent user group to provide iterative feedback, a literature review, a quantitative survey of adolescents, qualitative interviews with adolescents and parents, inductive thematic analysis of public reviews of VR exergames, a quantitative survey and qualitative interviews with users of the augmented reality running app Zombies, Run!, and building and testing a prototype with our adolescent user group.

Results: VR exergaming was appealing to adolescents and acceptable to parents. We identified behavior change techniques that users would engage with and features that should be incorporated into a VR exergame, including realistic body movements, accurate graphics, stepped levels of gameplay difficulty, new challenges, in-game rewards, multiplayer options, and the potential to link with real-world aspects such as physical activity trackers. We also identified some potential barriers to use, such as cost, perceived discomfort of VR headsets, and motion sickness concerns. A prototype game was developed and user-tested with generally positive feedback.

Conclusions: This is the first attempt to develop a VR exergame designed to engage adolescents in physical activity that has been developed within a public health intervention development framework. Our formative work suggests that this is a very promising avenue. The benefit of the design process was the collaborative parallel work between academics and game designers and the involvement of the target population in the game (intervention) design from the outset. Developing the game within an intervention framework allowed us to consider factors, such as parental support, that would be important for future implementation. This study also serves as a call to action for potential collaborators who may wish to join this endeavor for future phases and an example of how academic-industry collaboration can be successful and beneficial.

(J Med Internet Res 2021;23(2):e18161) doi:10.2196/18161

KEYWORDS

adolescent; adult; exercise; leisure activities; obesity; sports; video games; mobile phone; virtual reality; motivation

Introduction

Background

The benefits of performing sufficient physical activity (PA) are well established and include prevention of noncommunicable disease and better mental health [1,2]. Early adolescence (12-17 years) [3] is particularly important because there is a substantial age-related decline in activity levels from childhood into adolescence [4], and active adolescents are more likely to become active adults [5]. In addition, sustained moderate-to-vigorous PA (MVPA) in adolescence is positively associated with multiple markers of metabolic health, such as blood pressure levels, insulin, C-reactive protein, lipoprotein cholesterol, and triglycerides among others [6]. There is also some evidence that depression and anxiety rates are lower in more active adolescents [7], and PA reduces depressive symptoms in this group [8]. However, most of the adolescents are insufficiently active. Engaging adolescent girls in PA is particularly important because the age-related decline is greater for girls [9]. Objectively measured data from the Health Survey for England in 2016 showed that less than 15% of UK adolescents met the government's PA recommendations of at least 60 min MVPA per day [10]. Similarly, the United States Department of Health and Human Services estimates that over 80% of American adolescents do not meet guidelines [11], and low PA levels are observed globally [12,13]. Leisure-time PA throughout adolescence is increasingly replaced by sedentary behaviors such as screen time [14], highlighting screen time as a potential intervention target.

The Challenge of Engaging Adolescents in PA

Although it is clear that an intervention is required, how to change adolescent PA behavior at a population level remains unknown. A 2012 meta-analysis of 30 randomized controlled trials of PA interventions in children and adolescents up to 16 years of age found that interventions had negligible to small effects on accelerometer measured total activity levels and small effects of MVPA, equivalent to approximately 4 additional minutes per day. Only two of these interventions specifically targeted adolescents [15]. A 2019 meta-analysis of 17 school-based interventions, including children and adolescents, found no increase in objectively measured daily MVPA in the intervention group compared with control groups [16]. These authors suggested that interventions were failing at the implementation stage, and the paper was a call to action for a careful description of the intervention development and process evaluation [16]. A multimodal approach may be required to target school or home environments, policies, and parents [17,18]. However, multicomponent interventions are labor-intensive, and it remains unclear how feasible they are for wide-scale implementation. Digital interventions have been proposed as a solution, but their efficacy for promoting PA in adolescents is not clear. A systematic review identified 17 digital interventions for adolescents designed to promote PA and diet, but studies were generally small and had mixed findings [19]. In addition, the majority were web-based, which may not reflect current volitional adolescent technology behavior [19].

Gaming to Engage Adolescents in PA

Some industries have been very successful in engaging adolescents, particularly the gaming industry. Gaming is a recreational activity regularly performed by over 90% of adolescents, popular across socioeconomic groups and with girls and boys [20]. Gaming was recently highlighted as a promising avenue for health promotion [21]. Games requiring bodily movement—exergames—can encourage PA. A meta-analysis of 35 trials in children and adolescents found that previous generation exergames such as Wii Fit and Dance Dance Revolution had a similar physiological benefit to *active* controls (running or field-based PA) [22]. Exergames also enhanced self-efficacy, liking or enjoyment, and intrinsic motivation for PA [22]. In a 3-arm randomized controlled trial of participants aged between 15 and 19 years with obesity, those encouraged to play Wii Fit for 40-60 min per day cooperatively (working with a peer to expend calories) lost significantly more weight and had increased self-efficacy compared with competitive (competing against each other) or control conditions (regular daily activities) [23]. In another small trial by the same researchers, the percentage of body fat was reduced in girls who adhered to a dance exergame [24]. Although most exergames were not designed in a research context, McBain et al [25] developed a high-intensity training (HIT) exergame for adult males in deprived communities and found that the game delivered sustained progressive HIT training over a 6-week pilot [25]. A mobile app-based game called *Pokémon Go*, which has activity as a byproduct of the gaming and fun element, has had huge commercial success—with over 800 million downloads—and resulted in short-term increases in users' step count [26]. A longitudinal survey found that exergaming may be effective in engaging adolescent girls in sustained PA and that gender and motivation may be particularly important to consider when designing exergames [27].

The gaming industry has been commercially successful in building engaging products, but these were not developed as public health interventions or evaluated using traditional academic approaches. Collaboration between sectors offers a solution but can pose challenges, with industry and academia having different skills, goals, timeframes, epistemic cultures, and definitions of success [28]. The challenges of cross-sectoral collaboration between industry and academia and health services have long been recognized as barriers to innovation adoption in health care [20]. There are also differences within academia between researchers from a health background and those from a computing background. Blandford et al [29] note the contrast between the former having a focus on summative evaluation and the latter focusing on formative evaluation, as well as the differences in epistemology and reporting culture [29]. Academic computer science research is closer to software industry practice in a number of ways, such as the nature of filling the gaps in research (summative vs formative reviews), the accepted evidence, and the way research is reported [29].

Potential of Virtual Reality

As part of a previous small industry-partnered innovation grant designed to explore the potential of virtual reality (VR) to educate young people about the benefits of PA, *Innerselfie* [30], we conducted technology workshops with adolescents aged between 13 and 17 years allowing them to try top range virtual and augmented reality (AR) and observed how engaging they found immersive VR. VR has the potential to enhance the impact of an exergaming experience, as VR creates presence and allows the user to be fully immersed in the virtual environment [31]. VR headsets are increasingly portable, and wireless headsets and tracking allow for whole-body movement. Statista reported projected worldwide shipments of VR headsets to reach 15 million in 2022 [32]. There is early evidence from laboratory studies in adults that immersive VR exergaming is more engaging than standard exercise. For example, in 88 university staff and students, heart rate was higher in VR versus standard exercise conditions, with lower fatigue ratings and higher enjoyment in VR [33]. A pre-post study including 9 children found that playing an older generation VR exergame, *Astrojumper*, for 15 min enhanced self-reported motivation to exercise, but this study had no follow-up to test whether behavior changed [34]. However, to the best of our knowledge, when we began this work, there were no VR exergaming interventions designed to promote PA in adolescents and none where substantial formative development work had been conducted to inform content.

For the work presented in this paper, we followed the Medical Research Council (MRC) framework for developing and evaluating complex interventions published in 2000 and updated it in 2019 [35]. MRC guidance consists of 5 steps, (1) developing, (2) piloting, (3) evaluation, (4) reporting, and (5) implementation, specific to epidemiology, public health, and even social policy [35]. In line with the MRC framework of developing interventions and other coherent approaches to intervention development [35], substantial formative development work is recommended to understand the target behavior and end users' views. Although the interventions will likely need to target multiple social and environmental determinants of PA to achieve sustained PA change [36], we hypothesized that VR exergaming could be a platform to engage and motivate adolescents to be physically active and could ultimately form the core of a multidimensional intervention. However, it was recognized that to make a high-quality, appealing VR exergame that adolescents would want to play, industry experts would have to lead the development. Here, we describe the steps involved in the early intervention development process of a true industry-academic partnership aimed at designing a VR exergaming intervention for adolescents.

Objectives

The overarching objective of the vEngage project is to develop and test a PA intervention for younger adolescents involving VR exergaming. The purpose of this paper is to describe the process of the academic-industry partnership, formative development studies that fed into game development, and user testing of the prototype VR game. Future phases (subject to funding) will involve developing the game to full specification and identifying additional components (including *real-world* activity partners) to encourage sustained behavior change.

Funding was awarded by the UK MRC Public Health Intervention Development scheme to conduct the formative stages of intervention development outlined in Table 1. The project was a true academic-industry partnership whereby the industry partner Six to Start (led by AH) was a coinvestigator on the grant and involved since the project's inception; the grant gave structure to the collaboration. The collaboration was only possible because of the prior *problematization* [37] based on Six to Start's previous experience with the National Health Service, interest in the promotion of PA, and establishment of relationships across the research team (AH had been on the Innerselfie steering committee as an industry expert).

The academic team was led by AF, a behavioral scientist with a background in promoting PA for the prevention and treatment of chronic diseases. Other academic partners have expertise in sports science and epidemiology (LS), health informatics and statistics (HP), psychology and epidemiology (AS), risk perception and health behavior (KN), and psychology (NF). The industry partners (led by AH) are Six to Start [38], commercial game designers who developed the world's most successful mobile app AR narrative (audio)–based exergame Zombies, Run!

Table 1. Steps in phase I vEngage intervention development work.

Academic team	Update meeting	Industry team
Completed steps: phase I		
MRC[a] step 1: developing		
1. Literature review to identify determinants of PA[b] and mechanisms of change	X[c]	Scoping and finalizing suitable technologies
MRC step 2: piloting		
2. Recruit user group	X	Set up development environment and toolchain
3. Pilot quantitative survey with 695 adolescents	X	Preconception and technology feasibility
4. Qualitative interviews with adolescents and parents	X	Preconception and technology feasibility
5. Thematic analysis of public reviews of VR[d] exergames	X	Preconception
6. Survey of Zombies, Run! users with qualitative interviews	X	Develop and iterate prototype game using real hardware and exercise integration
Future steps: phase II		
MRC step 3: evaluation and MRC step 4: reporting		
MRC step 5: implementation		
7. User/steering committee testing	X	Prototype released to the research team
8. Pre-post study	—[e]	Not applicable for the industry team
9. Outline components for community PA link up and parental support	—	Outline necessary iterations to prototype game and possible link up technologies to integrate into game (eg, smartphone tracking)

[a]MRC: Medical Research Council.

[b]PA: physical activity.

[c]Steps completed by both teams.

[d]VR: virtual reality.

[e]Data not available.

Methods

Process for the Academic-Industry Collaboration

An important aspect of our intervention development was managing the academic-industry collaboration. It was essential to establish a process that could balance the need for formative development work using traditional research processes with the speed of working of game designers in a fast-moving technology-based industry within the period of grant funding for phase I (12 months). Ideally, formative work would have been conducted before game development, but this was not feasible within the funding period, so work had to be conducted in parallel, with ongoing dialog and regular meetings involving presentations of research findings (academic team) and game demonstrations (industry partners). Key steps and timings of meetings are shown in Table 1. The industry partner was always aware of the need for flexibility to be built into the game design in case new evidence arose from our work or the work of other researchers.

Literature Review: Identifying Theoretical and Behavioral Constructs

A key *first step* in the research process was to identify which determinants may influence younger adolescents' PA and therefore which theory may inform the intervention. NF extensively reviewed the literature and identified potentially modifiable targets (summarized in Table 2). The determinants identified most closely aligned with the Self-Determination Theory (SDT) of motivation [39], and it has been posited that the appeal of gaming may be grounded in their ability to satisfy basic psychological needs for competence, autonomy, and relatedness, core constructs of SDT [40]. The identified determinants and SDT were presented to the industry partners in our first formal meeting; then, the group discussed which would be feasible to target using the exergame component. As highlighted in the introduction, there is meta-analytic evidence that older generation exergaming can enhance PA self-efficacy, intrinsic motivation, and enjoyment in children and adolescents [22,41]. A pre-post VR exergame study in 9 children suggested VR exergaming may enhance motivation and enjoyment of PA [34]. Therefore, it was determined that a VR exergame could target PA self-efficacy, motivation, and identity by allowing young people to experience movement in an immersive and fun environment (Table 2). Determining the wider intervention components (Table 2) is a key part of phase II (future work, subject to funding).

Table 2. Behavioral determinants of adolescent physical activity from literature reviews and hypothesized method of inclusion in the intervention.

Potentially modifiable targets and description	Proposed intervention components	Proposed BCTs[a] [42]
VR[b] exergame[c]		
Self-efficacy: individuals own beliefs about their capability to carry out a task	Game features that encourage feedback on performance, small in-game rewards, and multiplayer elements for social support/feedback on the performance of others	• Feedback on performance • Goal setting • Rewards (in-game) • Social support • Time management
Motivation: particularly intrinsic motivation driven by internal reward	Delivery of PA[d] using a platform that is highly appealing to target users, perform PA in a fun and visually appealing immersive environment, and specific game targets to access next levels	• Goal setting • Self-reward • Incentive
PA identity: individuals view themselves as someone who is physically active (or not)	PA as a byproduct of fun and enjoyment, immersion as a distraction from negative physiological effects of exertion, and shifting negative perceptions of PA	• Framing and reframing • Identity associated with changed behavior
Wider intervention[e]		
Parental support: identify ways to reduce concerns about gaming and garner parental support for intervention	Design a nonviolent game; Displace sedentary with active screen time	• Social support • Monitoring of behavior by others without feedback (awareness) • Time management
Community/external PA opportunities: link up with PA partners that appeal to target users	User group suggestions were web-based influencers and users of trampolines, skate parks, or climbing walls	• Restructuring the physical and social environment • Behavioral practice or rehearsal
Link up with wearables and trackers: synchronize game with wearables or PA trackers to feed into the game	Most common or preferred by target group smartphone app	• Self-monitoring of behavior • Feedback on behavior
Habits	Encourage game play at the same time and the same context to develop a PA habit; Displacing sedentary behaviors with active gaming	• Habit formation • Planning

[a]BCT: behavior change technique.

[b]VR: virtual reality.

[c]Focus of the current phase (although some aspects like multiplayer could not be built into the prototype).

[d]PA: physical activity.

[e]Future work, subject to funding.

Behavior Change Techniques

The intervention required specific behavioral targets. It was not entirely possible to determine these from adolescent literature, since (as highlighted in the introduction) there is a lack of effective interventions to draw from. However, retrospective coding of more than 200 trials of PA interventions in more than 12,000 participants using behavior change techniques (BCTs) [43] found that BCT taxonomies such as goal setting, self-monitoring, and feedback on behavior were consistently associated with successful PA change [42,44]. The aforementioned review of digital interventions for adolescents also suggested that these BCTs may be particularly important [45]. However, part of the work within this grant (user group visits and quantitative survey with adolescents) aimed to determine which BCTs they would engage with in a digital intervention. Table 2 includes the proposed BCTs that could be linked to specific elements of the intervention.

Establishing and Working With a User Group

A key part of our formative work was establishing a group of adolescent users to obtain iterative feedback on ideas, help us design questions for empirical research, and try technologies. It was important to ensure representation from girls and boys. A total of 36 participants aged 13-16 years (19 boys and 17 girls) from 3 London schools were recruited. The user group was separated from the research participants. A number of visits to schools were conducted to (1) discuss initial ideas to shape the grant application, (2) discuss experiences of gaming and VR and rank BCTs in order of preference, and (3) try our prototype game and consider what the wider intervention might contain.

Quantitative Survey With Adolescents

A cross-sectional quantitative survey of adolescents was conducted through 2 London-based schools and web-based platforms (girls' schools were recruited in the hope of

understanding more about the views of girls, but the web-based version was open to anyone aged 13-24 years). The survey aimed to understand adolescent PA behavior, beliefs about and desire to change PA, and identify which BCTs they would like to see in a digital intervention. The survey also asked adolescents to report which platforms they used or preferred to help understand which companion technologies might accompany a VR exergame (to allow self-monitoring). Full analyses are in progress, but the summary findings using descriptive statistics (provided to the industry partner to incorporate into the game) are described in this paper.

Qualitative Interviews With Adolescents and Parents

Semistructured interviews were conducted with target users (adolescents aged 13-17 years) to explore their interest in VR, views about VR as a way to encourage activity, potential barriers, and which features they would like to see in a VR game. Interviews were conducted by 2 graduate psychology researchers. Data were analyzed using thematic analyses conducted by the academic team (AF, NF, and KN). To incorporate findings into game development, NF presented results as they emerged and created a table of desired game features. Full thematic analyses were performed [46,47] in parallel, fulfilling sufficient sample size requirements [48], and the findings were published (including the table of recommendations provided to the industry partner) in *JMIR Serious Games* in 2019 [49].

A separate qualitative study aimed to understand how parents of adolescents felt about VR as a way to encourage adolescent PA and how best to gain their support for this type of intervention. Summary findings necessary for game development were presented to the industry partners in a meeting, and full findings were published in *JMIR Serious Games* [50].

Inductive Thematic Analysis of Public Reviews of VR Exergames

To investigate game features that were particularly enjoyed or disliked in VR exergames reported players themselves, we thematically analyzed 498 publicly available reviews of exergames from the top 3 VR marketplaces Steam (Valve Corporation), Viveport (Valve Corporation), and Oculus (Oculus VR). A table of key recommendations was generated for the industry partner as soon as possible to share findings to feed into our game development, and the full study (including the table) has been published in *Journal of Medical Internet Research* [51].

Survey and Interviews With Zombies, Run! Users

The exergame developed by our industry partner (Zombies, Run) is an immersive audio AR mobile app released in 2012 [52]. The app became the highest-grossing health app on Apple's App Store 2 weeks post release. Since 2012, Zombies, Run! has accrued 5.5 million downloads, with approximately 200,000 monthly active users [53]. Zombies, Run! combines exercise and a postapocalyptic radio story, a narrative-based game delivered via the smartphone app. To understand what appealed to the users of popular and widely used exergames and their potential impact on PA behavior, we surveyed Zombies, Run! users between November and January 2019. The survey was cross-sectional and included 36 questions around experiences of using Zombies, Run!, likes and dislikes of app features, and engagement with BCTs. Users were also asked to report their PA levels (number of days and length of sessions of MVPA using the items from Coleman et al [54]) before and after using Zombies, Run! The survey also aimed to explore their interest in VR exergaming and perceived benefits and barriers to this. The users were reached via in-app notifications, Zombies, Run! social media, and Zombies, Run! newsletter. The survey asked Zombies, Run! users if they were willing to be interviewed qualitatively to tell us more about their experiences, and 30 Zombies, Run! users were interviewed.

Building the Prototype

At the conception of the project, the industry partners developed gameplay concepts for the VR prototype based on some key findings that emerged from our research (summarized in Table 3). Six to Start then established some basic principles to widen the potential game audience in terms of technology and appeal:

1. The game should not require any special equipment other than a standard Vive/Oculus VR setup, that is, no weights, additional Vive trackers, exercise bikes, or pull-up bars (because this would limit future implementation).
2. Gameplay should not be overtly about fitness or exercise (because our user group almost universally reported they would prefer a fun game with physical movement as a byproduct).
3. The game must work in a standard household environment, such as a living room with a ceiling with normal height.

Six to Start reviewed a wide set of basic gameplay types, such as Simon Says (similar to Dance Dance Revolution), Dodgeball, Point and Shoot, Building, Plate-spinning (maintenance of a chaotic system, eg, Diner Dash), and Photography. They also established some desirable principles such as quick sessions, peaks and troughs in physical exertion, and quick start-up, all of which are common in very popular *casual* and *hypercasual* games that tend to reach a very wide audience. Finally, both teams discussed the content of the game and how it might be scaled—whether the game's levels should be all human-authored or randomized or semirandomized. Two broad concepts were arrived at with the working titles *Action Photography* and *Hole in the Wall*. These were discussed with our user group, ultimately to develop solely the latter concept to have more time for iteration and graphical polish. Development took place on the Unity platform and Steam VR to ease cross-platform deployment to different VR hardware. The prototype was designed for room-scale use and tested on HTC Vive hardware with the standard 2 hand controllers to allow for full-body movement tracking. Gameplay was iterated during development, adding music, and developed levels that increased in difficulty and built level design tools for nonprogrammers.

Table 3. Key features of research fed into the game design.

Study	Key findings that fed into game design	Reference
Review of the literature	• Presented in Table 2	Table 2
User group	• Preferences for game types and preferred BCTs[a] • Iterative feedback on prototype ideas	N/A[b]
Adolescent survey	• Strong desire to increase PA[c] • Preferred BCTs such as goal setting and feedback on performance	In preparation
Adolescent qualitative interviews	• Desired VR[d] exergame features: being able to exercise at home; rewards; increasing challenges; social or multiplayer aspects; using own music • Barriers: high cost; the need for parental buy-in	Published in a study by Farič et al, 2019a [49]
Parent qualitative interviews	• Approval of harnessing gaming for something positive • Intervention must be nonviolent • Mental health most salient reason for encouraging PA, and a calming virtual environment preferred	Published in a study by McMichael et al 2020 [50]
Thematic review of VR exergame reviews	• Desired VR exergame features: removing motion sickness owing to the immersive quality and accurate graphics; gradual acquisition of skill and multiplayer options with music • Disliked features: motion sickness, poor graphics, and unresponsive developers	Published in a study by Farič et al, 2019b [51]
ZR[e] survey	• Global appeal to those who identified as gamers and nongamers of all ages; primary use of ZR was running but also walking, gardening, cycling, and training for weight and fitness control, even to run marathons	In preparation
ZR interviews	• The narrative and interactive storyline as the reason for engagement, not the actual gameplay; a distraction from negative and mundane aspects of running; community feel; more than half used ZR to improve mental health (improved mood)	In preparation

[a]BCT: behavior change technique.

[b]N/A: not applicable.

[c]PA: physical activity.

[d]VR: virtual reality.

[e]ZR: *Zombies, Run!*

Intellectual Property: Game Code

Establishing intellectual property (IP) agreements with multiple partners can be complex, and establishing an IP contract that satisfies all collaborators took many months. A funder requirement was that, at the end of the project, the code for our game would be made available open source under the Open Source License GPL v3 (General Public License version 3) and available on our website. Open-source code is beneficial for future software development and the academic community. A benefit of acquiring external funding for a research endeavor meant that Six to Start were also involved in an exploratory research capacity, rather than for financial gain. If other partnerships requiring similar IP contracts wish to see the content of the agreement to expedite their own process, they are welcome to contact the corresponding author for sharing.

Ethical Consideration

The University College London (UCL) Ethics Committee provided ethical approval for all relevant steps described above (Project IDs: 10213/001, 12669/001, and 3777/004) and adolescents or parents provided informed written consent. The results of each stage of this work have been or will be disseminated through presentations at conferences in PA, public health, and gaming, and peer-reviewed publications.

Results

Quantitative Survey of Adolescents

A total of 511 adolescents from schools and youth organizations completed the survey. Of these, 48.9% (250/511) were aged 13-15 years, 35.6% (182/511) aged 16-18 years and 15.5% (79/511) were aged 19-24 years. Analyses are ongoing, and we plan to explore differences in preferred BCTs and technology features by gender and age; however, 77.1% (394/511) were interested in advice to increase their PA and most likely to go on the web or use a tablet or mobile app to find information about PA. About 50.0% (256/511) used some kind of health tracker. Preliminary data on preferred BCTs suggest that goal

setting, personalized feedback on behavior, instructions on how to perform a behavior, self-monitoring, and rewards were strongly desired. Features that were less desired were information about the health consequences of not performing sufficient PA and features that included social networks or forums or photo feedback.

Qualitative Interviews With Adolescents and Parents

Qualitative interviews were conducted with 31 adolescents (18/31, 58% female; 19/31, 62% non-White ethnicities). Boys and girls were equally positive about the use of VR for PA promotion. Both highlighted fun or enjoyment as fundamental. Participants identified rewards, increasing challenges, and including social or multiplayer and real-world aspects as important game features. Barriers included the cost of high-end systems and the need for parental approval [49]. Exercising at home was perceived as very appealing and a way to overcome social and cultural barriers to PA, particularly for girls. A total of 18 parents of adolescents took part in interviews and believed that VR exergaming would engage their adolescents with PA, and although they would prefer real-world PA, they were very supportive of an intervention that harnessed gaming for a positive outcome (promotion of PA). In addition, they consistently reported that to garner their support, a game must be nonviolent, and that mental health was their most salient reason for encouraging PA in their adolescents, so a game would ideally be calming or relaxing rather than aggressive. For more details on the results of the studies, see studies by Farič et al [49] and McMichael et al [50].

Thematic Analysis of Public Reviews of VR Exergames

The results of the review found that VR exergaming was a way to engage with PA in a fun, enjoyable, and playful way, without PA being the focus of the activity. Promisingly, users reported feeling that VR exergaming had provided exertion comparable with *real-world* PA. However, some notable drawbacks were also identified, such as those pertaining to the technology itself (eg, the mechanics of the games and unintuitive controls) and a lack of real-world feeling while playing the games. The full details of this study are available in a study by Farič et al [51].

Survey of Zombies, Run! Users

This work was in progress at the time of writing, and we plan to explore findings by age and gender, but summary findings are presented here. A total of 6423 participants opened the link, and 83.19% (5343/6423) completed the survey. Participants' age ranged from 16 to 71 years (mean 33.1, SD 10.1), with 58.62% (3131/5341) of them identified as female, 38.04% (2032/5341) as male, and 2.52% (135/5341) as other. Most participants were White (4498/5341, 84.21%), and approximately half were from the United States (2594/5343, 48.54%). The most common education category was a bachelors' degree education (1979/5341, 37.05%). Zombies, Run! app usage was associated with a reported mean increase of 84 min in PA per week (95% CI 82-87). Overall, 39.89% (2119/5311) of participants experienced a positive identity shift (from not a runner to a runner). The BCTs or game features with the strongest perceived impact on the PA behavior were positive outcomes of PA, goal setting, and obtaining intrinsic reward (through fun), whereas the least important were obtaining reward money and social comparison. The most popular game features of Zombies, Run! included simulation, customization, self-monitoring, roleplay, and obtaining a within-game reward. A total of 57.92% (3095/5343) had not tried VR exergaming but would like to, 20.11% (1075/5343) had not tried it and did not want to, 7.44% (398/5343) had tried and liked it, 1.59% (85/5343) had tried and did not like it, and 1.94% (104/5343) did not know what it was. By far, the most appealing perceived positive aspects of VR were immersion and fun or enjoyment. The perceived negative aspects were cost and discomfort (heavy or bulky headset).

Preliminary thematic analysis of 30 interviews revealed that people became immersed in PA through the story's narrative, which motivated them to keep going and distracted them from negative associations with PA. The app was not solely used for running by all participants, but also for walking, gardening, or cycling. A total of 70% (21/30) of the interviewed sample reported positive effects on mental health. The qualitative and quantitative work is being written for publication by NF as part of her PhD work.

The Prototype

The key findings of the formative research fed into the prototype game are described briefly in Table 3. The game was given the working title *Walls* (although we subsequently asked the user group to suggest appealing names; Figure 1). In the game, the player needs to use the VR controllers to complete complex patterns that appear on walls by moving their arms and body to be as accurate as possible. An image of a player trying the prototype is presented in Figure 2. The accuracy and speed of completion earn players high scores. The game includes PA movements such as stretching, fast and slow arm motions, and dodging. The idea is that if games were developed further, additional levels would become physically challenging. However, the virtual environment was also designed to be calming (based on parental views).

Figure 1. Walls exergame developed and evaluated as part of the project.

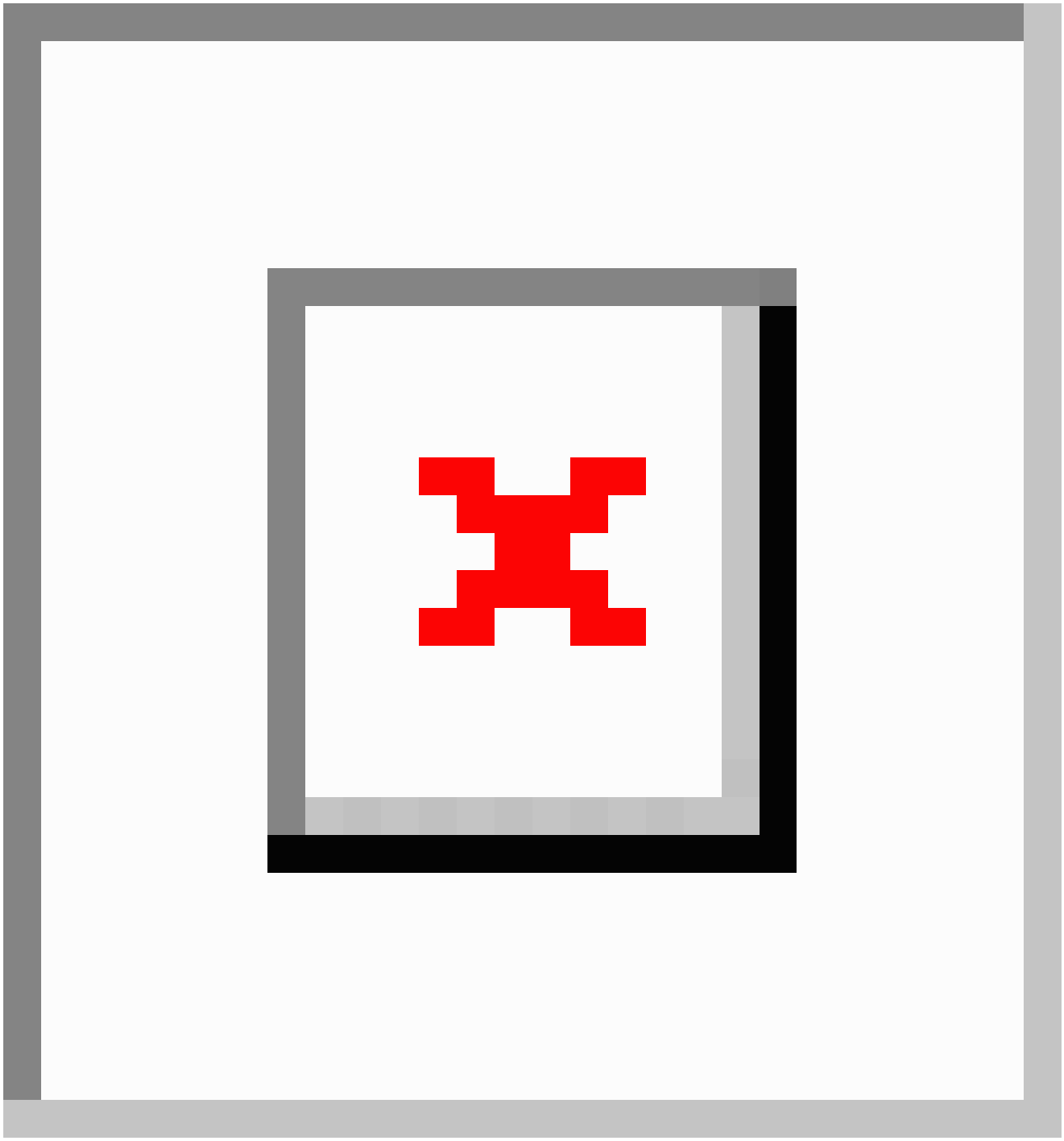


Figure 2. A player during Walls gameplay.

User Testing of the Prototype

Before trying our prototype, 89% (17/19) of the boys and 59% (10/17) of girls had tried some form of VR. This was almost always a smartphone-based headset rather than a fully immersive experience. All but one who had tried it reported enjoying it, except for one reporting motion sickness. When asked which features they would like to see in our game, goal setting, rewards, social aspects, choosing their own music, custom levels or custom storyline, player creation, and progressive game challenges were reported as the most important features. Most students (but not all) enjoyed the prototype gameplay (23/31, 74%). Issues were lack of instructions at the set up to practice, confusing controller functions, and not seeing the overview of other people's scores at the end to compare their performance. Almost all students reported feeling excited about the potential of the VR game once it had been further developed and provided additional feedback ranging from alternative names, features that would make the game even more appealing, to ideas for the wider intervention.

Other Important Outcomes: Training of Researchers and Engaging Adolescents With Science

The formative development work described here has facilitated substantial postgraduate training. In a short funding timeframe, when a large proportion of the grant has to necessarily go toward game development, to conduct and disseminate work, it has been invaluable to involve graduate psychology researchers. In total, 4 UCL Master of Science students (listed in the Acknowledgments sections in this paper and authors on the relevant papers) conducted their dissertations on this work, and the data will contribute to the PhD of NF (who was employed as a research assistant during the funding period). This project's novelty and the opportunity to spend time in a research center and a gaming lab made the projects appealing and extremely enjoyable and valuable for students. In addition, teachers from the schools where the user groups were recruited informed us that an unanticipated benefit was the engagement of their young people in the scientific process, with a number of their girls and boys expressing feeling motivated to pursue careers in scientific research or computer science or game development.

Discussion

Principal Findings

This paper described the formative intervention development work that resulted in the development of a prototype VR exergame designed to engage adolescents with PA. The results of the vEngage formative development phases demonstrate that our target population—younger adolescents aged between 13 and 17 years—see great potential in VR technology as a means of engaging in PA, and this enthusiasm was reflected by users of another exergame population (Zombies, Run!). We identified a number of desired features of VR exergames, including realistic body movements, removing motion sickness and accurate graphics, gradual acquisition of skills, and multiplayer options with music. However, notable perceived drawbacks to a VR exergaming intervention were affordability, accessibility, and potential discomfort. In addition, our study using inductive thematic analysis of public VR exergame reviews suggested that although VR exergaming can elicit levels of exertion that users equate with other forms of MVPA and distract participants from the negative perception of performing PA, the affordability of high-end VR equipment, graphics, and motion sickness are still drawbacks to VR becoming more mainstream [51].

The benefit of the applied design was that academics used traditional academic methods of inquiry to study the target population and, in turn, informed the game developers. Feedback on our prototype was encouraging, and future work (subject to

funding) will seek to develop the prototype and wider intervention components further. The academic-industry collaboration was successful because of its iterative nature and frequent meetings where goals were set and the work was performed in parallel and shared in real time. We hope that our process can guide researchers who wish to design an intervention together with their target population and an industry partner. We established ways of distilling top-line academic findings quickly to provide to an industry partner. In addition, we used methods of accessing information quickly, such as analyzing existing publicly available reviews of VR exergames [51] or collecting data from the existing Zombies, Run! user base, along with more traditional recruitment processes.

Study Strengths and Limitations

This study has several strengths and limitations. The collaborative approach and involvement of end users were strengths. However, within the limits of early-phase funding, it was only possible to develop the game to a standard that would facilitate user input, meaning that we could not assess the game's potential for impact on PA. We have an ongoing experimental study exploring the potential exertion (heart rate and perceived exertion) that can be achieved by playing some of the popular, commercially available VR games.

The mixed methods were a strength; we employed a number of different methods, including thematic analysis of player reviews, app surveys, standard surveys, and qualitative interviews. However, formative studies had limitations. The quantitative surveys were cross-sectional self-report (although questions about interest and engagement cannot be assessed objectively), and a future aim would be to test the fully developed game and use accelerometers to measure PA. The school-based survey was completed as a class activity following opt-out consent. However, it is very likely that the Zombies, Run! user survey was completed by users who felt most positive about the exergame, introducing some selection bias. Our user group and those interviewed qualitatively were recruited from schools in and around the research center in London, which may not reflect others' views in different geographical locations (although the sample of interviewed Zombies, Run! users was global). The vast majority of those we surveyed or interviewed did not own or use VR frequently, which meant that barriers and benefits were generally perceived rather than based on experience. Prolonged and frequent exposure in the context of a trial would allow us to explore how these factors influence their PA behavior and user experience. A multimodal intervention would likely be required, targeting multiple influences on adolescent PA [17,18], and this work only focused on VR exergaming (which we hypothesized may enhance motivation for PA). In line with frameworks such as the COM-B model (capability, opportunity, motivation, behavior) and behavior change wheel, it is important that an intervention considers the broader influences on behavior and not just individual motivation [42].

Long-Term, Sustained Behavior Change

It is worth considering the broader application of our findings, including steps toward implementation of VR (and AR) exergaming interventions. Implementation of VR games to sustain behavior change would be made easier if VR gyms or fitness centers using VR existed in the United Kingdom, but at present (August 2020), there are none, and the cost of an average head-mounted VR is still relatively high (average Can $585 [US $460]). The virtual gyms (eg, Digme) that do exist do not allow full immersion via head-mounted pieces but modify the user's experience by similar means such as theater performances and AR (eg, use of music and lights) and display certain features on the screens in front of users [55]. The availability and accessibility of VR for PA is still tied to cost, the technological market, and PA competitors (eg, gyms, fitness clubs, sports centers, apps, and outdoors). VR exergaming would have to offer people a distinct and accessible PA experience. However, it has been predicted that VR console homeownership will continue to rise [32]. The COVID-19 pandemic and the need for social distancing or restrictions placed on exercise facilities mean that tools that can facilitate home-based exercise have never been more important.

Future Work

There are a number of challenges in the next steps of the project. First, funding must be acquired for future phases. The funding was awarded for formative development work, and we believe that we have established that VR exergaming is promising for engagement with PA in our target group. However, significant development work is required for the game to be at the stage where it is ready to trial. As highlighted in the introduction, a multimodal intervention would likely be required to facilitate sustained PA change [17,18]. We have always acknowledged that the game alone is unlikely to be sufficient for long-term change and working with our user group identified that linking with *real-world* technologies (eg, trackers) and real-world activity partners (eg, vouchers to try a range of real-world activities like climbing wall, boxing class, and trampoline park). However, these options will have to consider the post–COVID-19 pandemic world and the need for socially distanced exercise, which our vEngage project did not directly address. The implementation also requires the game to be promoted in order to gain a sufficient user base to support further testing and warrant further development. To be successful as more than a testbed, the game needs a funding mechanism beyond the initial research funding.

Future intervention development steps (phase II) include developing the VR game to a high specification, identifying components that can link PA to the game (eg, smartphone phone tracking), linking with real-world PA partners (eg, schools and community or leisure centers, gaming conferences), and linking with vloggers (video bloggers). Ultimately, we would like to empirically test our intervention's potential to enhance motivation and change behavior (phase II).

Conclusions

This vEngage project is the first attempt to develop a VR exergame designed to engage adolescents with PA using an academic-industry collaboration [56]. Our findings suggest that VR exergaming has potential as a public health intervention designed to engage adolescents in PA. We plan to take this work forward and invite collaboration for future stages.

Acknowledgments

The authors would like to thank the UK MRC for funding the vEngage project for funding (grant number MR/R015430/1). The authors would like to acknowledge the contributions of 4 UCL MSc students, Lucy McMichael, James Botevyle, Eleanor Yorke, and Katie Rudge, to the described work, and the authors thank the students in their user group and the teachers who facilitated these.

Conflicts of Interest

As outlined in the paper, this is a true academic and industry collaboration funded by the UK MRC industry partnership grant and leads to the development of a VR game licensed by Six to Start. There is no legal, financial, or commercial conflict with the authors' industry partner company, Six to Start.

References

1. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
2. Blondell SJ, Hammersley-Mather R, Veerman JL. Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies. BMC Public Health 2014;14:510 [FREE Full text] [doi: 10.1186/1471-2458-14-510] [Medline: 24885250]
3. Hallal PC, Victora CG, Azevedo MR, Wells JCK. Adolescent physical activity and health: a systematic review. Sports Med 2006;36(12):1019-1030. [doi: 10.2165/00007256-200636120-00003] [Medline: 17123326]
4. Telama R, Yang X, Viikari J, Välimäki I, Wanne O, Raitakari O. Physical activity from childhood to adulthood: a 21-year tracking study. Am J Prev Med 2005 Apr;28(3):267-273. [doi: 10.1016/j.amepre.2004.12.003] [Medline: 15766614]
5. Bell JA, Hamer M, Richmond RC, Timpson NJ, Carslake D, Davey Smith G. Associations of device-measured physical activity across adolescence with metabolic traits: orospective cohort study. PLoS Med 2018 Sep 11;15(9):e1002649 [FREE Full text] [doi: 10.1371/journal.pmed.1002649] [Medline: 30204755]
6. Schmoldt A, Rohloff C. Dehydro-digitoxosides of digitoxigenin: formation and importance for the digitoxin metabolism in the rat. Naunyn Schmiedebergs Arch Pharmacol 1978 Nov;305(2):167-172. [doi: 10.1007/BF00508288] [Medline: 732892]
7. Bélair M, Kohen DE, Kingsbury M, Colman I. Relationship between leisure time physical activity, sedentary behaviour and symptoms of depression and anxiety: evidence from a population-based sample of Canadian adolescents. BMJ Open 2018 Oct 17;8(10):e021119 [FREE Full text] [doi: 10.1136/bmjopen-2017-021119] [Medline: 30337306]
8. Bailey AP, Hetrick SE, Rosenbaum S, Purcell R, Parker AG. Treating depression with physical activity in adolescents and young adults: a systematic review and meta-analysis of randomised controlled trials. Psychol Med 2018 May;48(7):1068-1083. [doi: 10.1017/S0033291717002653] [Medline: 28994355]
9. Dumith SC, Gigante DP, Domingues MR, Kohl HW. Physical activity change during adolescence: a systematic review and a pooled analysis. Int J Epidemiol 2011 Jun;40(3):685-698 [FREE Full text] [doi: 10.1093/ije/dyq272] [Medline: 21245072]
10. Health Survey for England 2016 Children's health. National Centre for Social Research, UCL. London: NHS Digital; 2017. URL: http://healthsurvey.hscic.gov.uk/media/63769/HSE2016-Child-health.pdf [accessed 2020-12-12]
11. Landry BW, Driscoll SW. Physical activity in children and adolescents. PM R 2012 Nov;4(11):826-832. [doi: 10.1016/j.pmrj.2012.09.585] [Medline: 23174545]
12. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. The Lancet Child & Adolescent Health 2020 Jan;4(1):23-35. [doi: 10.1016/s2352-4642(19)30323-2]
13. Xu J, Gao C. Physical activity guidelines for Chinese children and adolescents: the next essential step. J Sport Health Sci 2018 Jan;7(1):120-122 [FREE Full text] [doi: 10.1016/j.jshs.2017.07.001] [Medline: 30356441]
14. Corder K, Sharp SJ, Atkin AJ, Griffin SJ, Jones AP, Ekelund U, et al. Change in objectively measured physical activity during the transition to adolescence. Br J Sports Med 2014 Apr 9:730-736 [FREE Full text] [doi: 10.1136/bjsports-2013-093190] [Medline: 24273308]
15. Metcalf B, Henley W, Wilkin T. Effectiveness of intervention on physical activity of children: systematic review and meta-analysis of controlled trials with objectively measured outcomes (EarlyBird 54). BMJ 2012 Sep 27;345:e5888. [doi: 10.1136/bmj.e5888] [Medline: 23044984]
16. Love R, Adams J, van Sluijs EMF. Are school-based physical activity interventions effective and equitable? A meta-analysis of cluster randomized controlled trials with accelerometer-assessed activity. Obes Rev 2019 Jun;20(6):859-870 [FREE Full text] [doi: 10.1111/obr.12823] [Medline: 30628172]
17. Messing S, Rütten A, Abu-Omar K, Ungerer-Röhrich U, Goodwin L, Burlacu I, et al. How can physical activity be promoted among children and adolescents? A systematic review of reviews across settings. Front Public Health 2019 Mar 19;7:55-55 [FREE Full text] [doi: 10.3389/fpubh.2019.00055] [Medline: 30941342]

18. Sutherland R, Campbell E, Lubans DR, Morgan PJ, Okely AD, Nathan N, et al. 'Physical Activity 4 Everyone' school-based intervention to prevent decline in adolescent physical activity levels: 12 month (mid-intervention) report on a cluster randomised trial. Br J Sports Med 2016 Apr 10;50(8):488-495 [FREE Full text] [doi: 10.1136/bjsports-2014-094523] [Medline: 26359346]
19. Rose T, Barker M, Maria JC, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
20. Ii SS, Fitzgerald L, Morys-Carter MM, Davie NL, Barker R. Knowledge translation in tri-sectoral collaborations: an exploration of perceptions of academia, industry and healthcare collaborations in innovation adoption. Health Policy 2018 Feb;122(2):175-183. [doi: 10.1016/j.healthpol.2017.11.010] [Medline: 29191387]
21. Goodman W, McFerran E, Purves R, Redpath I, Beeken RJ. The untapped potential of the gaming community: narrative review. JMIR Serious Games 2018 Sep 25;6(3):e10161 [FREE Full text] [doi: 10.2196/10161] [Medline: 30274962]
22. Gao Z, Chen S, Pasco D, Pope Z. A meta-analysis of active video games on health outcomes among children and adolescents. Obes Rev 2015 Sep;16(9):783-794. [doi: 10.1111/obr.12287] [Medline: 25943852]
23. Staiano AE, Abraham AA, Calvert SL. Adolescent exergame play for weight loss and psychosocial improvement: a controlled physical activity intervention. Obesity (Silver Spring) 2013 Mar;21(3):598-601 [FREE Full text] [doi: 10.1002/oby.20282] [Medline: 23592669]
24. Staiano AE, Marker AM, Beyl RA, Hsia DS, Katzmarzyk PT, Newton RL. A randomized controlled trial of dance exergaming for exercise training in overweight and obese adolescent girls. Pediatr Obes 2017 Apr;12(2):120-128 [FREE Full text] [doi: 10.1111/ijpo.12117] [Medline: 26918815]
25. McBain T, Weston M, Crawshaw P, Haighton C, Spears I. Development of an Exergame to Deliver a Sustained Dose of High-Intensity Training: Formative Pilot Randomized Trial. JMIR Serious Games 2018 Mar 27;6(1):e4 [FREE Full text] [doi: 10.2196/games.7758] [Medline: 29588271]
26. Howe KB, Suharlim C, Ueda P, Howe D, Kawachi I, Rimm EB. Gotta catch'em all! Pokémon GO and physical activity among young adults: difference in differences study. BMJ 2016 Dec 13;355:i6270 [FREE Full text] [doi: 10.1136/bmj.i6270] [Medline: 27965211]
27. O'Loughlin EK, Barnett TA, McGrath JJ, Consalvo M, Kakinami L. Factors associated with sustained exergaming: longitudinal investigation. JMIR Serious Games 2019 Jul 31;7(2):e13335 [FREE Full text] [doi: 10.2196/13335] [Medline: 31368440]
28. Ankrah S, AL-Tabbaa O. Universities–industry collaboration: a systematic review. Scandinavian Journal of Management 2015 Sep;31(3):387-408. [doi: 10.1016/j.scaman.2015.02.003]
29. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
30. Fisher A, Meisel S, Webb T, Fisher B. Innerselfie: Can we use Immersive Virtual Reality to change cancer risk perceptions and promote physical activity in young people? In: UCL 2nd Behaviour Change Conference: Digital Health and Wellbeing. SENATE HOUSE, MALET STREET, LONDON, WC1E 7HU: Frontiers of Public Health; 2016:2.
31. Slater M, Wilbur S. A Framework for Immersive Virtual Environments (FIVE): speculations on the role of presence in virtual environments. Presence: Teleoperators & Virtual Environments 1997 Dec;6(6):603-616. [doi: 10.1162/pres.1997.6.6.603]
32. Statista. 2020. Worldwide VR headset shipment 2018-2022, by segment. 2021 Jan 4. URL: https://www.statista.com/statistics/754860/worldwide-vr-headset-shipment-by-segment/ [accessed 2021-01-07]
33. Plante TG, Aldridge A, Bogden R, Hanelin C. Might virtual reality promote the mood benefits of exercise? Computers in Human Behavior 2003 Jul;19(4):495-509. [doi: 10.1016/S0747-5632(02)00074-2]
34. Finkelstein S, Suma EA. Astrojumper: motivating exercise with an immersive virtual reality exergame. Presence: Teleoperators and Virtual Environments 2011 Feb 01;20(1):78-92. [doi: 10.1162/pres_a_00036]
35. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337(sep29 1):a1655-a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
36. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJF, Martin BW, et al. Correlates of physical activity: why are some people physically active and others not? Lancet 2012 Jul 21;380(9838):258-271. [doi: 10.1016/S0140-6736(12)60735-1] [Medline: 22818938]
37. Callon M. Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. The Sociological Review 2015 May 29;32(1_suppl):196-233. [doi: 10.1111/j.1467-954x.1984.tb00113.x]
38. Six to Start. Six to Start. URL: https://www.sixtostart.com/ [accessed 2020-06-16]
39. Theory. Center for Self Determination Theory. URL: https://selfdeterminationtheory.org/theory/[accessed [accessed 2021-01-05]
40. Przybylski AK, Rigby CS, Ryan RM. A motivational model of video game engagement. Rev of Gen Psychol 2010 Jun;14(2):154-166. [doi: 10.1037/a0019440]

41. Andrade A, Correia CK, Coimbra DR. The psychological effects of exergames for children and adolescents with obesity: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2019 Nov;22(11):724-735. [doi: 10.1089/cyber.2019.0341] [Medline: 31697604]
42. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
43. Michie S, van SMM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
44. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Dec 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
45. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
46. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
47. Clarke V, Braun V. Successful qualitative research: a practical guide for beginners. In: Research Methods in Psychology. Thousand Oaks, California, United States: SAGE Publications Ltd; 2013:1-400.
48. Fugard AJ, Potts HW. Supporting thinking on sample sizes for thematic analyses: a quantitative tool. Int J of Soc Res Method 2015 Feb 10;18(6):669-684. [doi: 10.1080/13645579.2015.1005453]
49. Faric N, Yorke E, Varnes L, Newby K, Potts HW, Smith L, et al. Younger adolescents' perceptions of physical activity, exergaming, and virtual reality: qualitative intervention development study. JMIR Serious Games 2019 Jun 17;7(2):e11960 [FREE Full text] [doi: 10.2196/11960] [Medline: 31210135]
50. McMichael L, Faric N, Newby K, Potts H, Hon A, Smith L, et al. Physical activity, gaming and virtual reality: the views of parents of adolescents. JMIR Serious Games 2020;8(3):e14920. [doi: 10.2196/preprints.14920] [Medline: 32840487]
51. Faric N, Potts HWW, Hon A, Smith L, Newby K, Steptoe A, et al. What players of virtual reality exercise games want: thematic analysis of web-based reviews. J Med Internet Res 2019 Sep 16;21(9):e13833 [FREE Full text] [doi: 10.2196/13833] [Medline: 31538951]
52. Getting fit with the undead: the Zombies, run story. Alphr Business. online: Alphr; 2017 May 15. URL: https://www.alphr.com/business/1005907/getting-fit-with-the-undead-the-zombies-run-story [accessed 2020-05-06]
53. Zombies, run!. App Store Preview. URL: https://apps.apple.com/gb/app/zombies-run/id503519713 [accessed 2020-05-06]
54. Coleman KJ, Ngor E, Reynolds K, Quinn VP, Koebnick C, Young DR, et al. Initial validation of an exercise. Med Sci Sports Exerc 2012 Nov;44(11):2071-2076. [doi: 10.1249/MSS.0b013e3182630ec1] [Medline: 22688832]
55. Workouts for everybody. DIGME. URL: https://www.digmefitness.com/ [accessed 2020-06-06]
56. vEngage study. vengagestudy. 2018 Nov 11. URL: https://vengagestudy.com [accessed 2020-06-16]

Abbreviations

AR: augmented reality
BCT: behavior change technique
HIT: high-intensity training
IP: intellectual property
MRC: Medical Research Council
MVPA: moderate-to-vigorous physical activity
PA: physical activity
SDT: Self-Determination Theory
UCL: University College London
VR: virtual reality

Edited by G Eysenbach; submitted 07.02.20; peer-reviewed by T Muto, E O'Loughlin, K Alipasandi; comments to author 23.03.20; revised version received 28.06.20; accepted 11.11.20; published 04.02.21.

Please cite as:
Farič N, Smith L, Hon A, Potts HWW, Newby K, Steptoe A, Fisher A
A Virtual Reality Exergame to Engage Adolescents in Physical Activity: Mixed Methods Study Describing the Formative Intervention Development Process
J Med Internet Res 2021;23(2):e18161
URL: http://www.jmir.org/2021/2/e18161/
doi:10.2196/18161
PMID:33538697

Original Paper

Identifying Self-Management Support Needs for Pregnant Women With Opioid Misuse in Online Health Communities: Mixed Methods Analysis of Web Posts

Ou Stella Liang[1], MHA; Yunan Chen[2], PhD; David S Bennett[3], PhD; Christopher C Yang[1], PhD

[1]College of Computing and Informatics, Drexel University, Philadelphia, PA, United States

[2]Donald Bren School of Information and Computer Sciences, University of California, Irvine, Irvine, CA, United States

[3]Department of Psychiatry, College of Medicine, Drexel University, Philadelphia, PA, United States

Corresponding Author:
Christopher C Yang, PhD
College of Computing and Informatics
Drexel University
3675 Market St
Philadelphia, PA
United States
Phone: 1 215 895 1631
Email: chris.yang@drexel.edu

Abstract

Background: The current opioid crisis in the United States impacts broad population groups, including pregnant women. Opioid use during pregnancy can affect the health and wellness of both mothers and their infants. Understanding women's efforts to self-manage opioid use or misuse in pregnancy is needed to identify intervention points for improving maternal outcomes.

Objective: This study aims to identify the characteristics of women in an online health community (OHC) with opioid use or misuse during pregnancy and the self-management support needs of these mothers.

Methods: A total of 200 web posts by pregnant women with opioid use participating in an OHC were double coded. Concepts and their thematic connections were identified through an inductive process until theoretical saturation was reached. Statistical tests were performed to identify patterns.

Results: The majority of pregnant women (150/200, 75.0%) in the OHC exhibited signs of misuse, and 62.5% (125/200) of the participants were either contemplating or pursuing dosage reduction. Self-managed withdrawal was more common (P<.001) than professional treatment among the population. A total of 5 themes of self-management support needs were identified as women sought information about the potential adverse effects of gestational opioid use, protocols for self-managed withdrawal, pain management safety during pregnancy, hospital policies and legal procedures related to child protection, and strategies for navigating offline support systems. In addition, 58.5% (117/200) of the pregnant women expressed negative emotions, of whom only 10.2% (12/117) sought to address their emotional needs with the help of the OHC.

Conclusions: OHCs provide vital self-management support for pregnant women with opioid use or misuse. Women pursuing self-managed dosage reduction are prone to misinformation and repeated relapses, which can result in extreme measures to avoid testing positive for drug use at labor. The study findings provide evidence for public policy considerations, including universal screening of substance use for pregnant women, emphasis on treatment rather than legal punishment, and further expansion of the Drug Addiction Treatment Act waiver training program. The improvement of web-based platforms that can organize geo-relevant information, dispense clinically validated withdrawal schedules, and offer structured peer support is envisioned for harm reduction among pregnant women who opt for self-management of opioid misuse.

(J Med Internet Res 2021;23(2):e18296) doi:10.2196/18296

KEYWORDS

self-management; online health community; opioid use disorder; pregnancy

Introduction

Background

> *I literally feel like this is the end. Like there is no way out. ...I don't want the heroin to harm the baby and I don't want withdrawals to harm the baby. And I don't have the money...* [P1311]

This is a quote from a pregnant woman in an online health community (OHC) who shared her dire circumstances. She is one of an estimated 21,000 pregnant women who reported misuse of opioids in the past month in the United States [1]. In a recent Centers for Disease Control and Prevention study, 6.6% of women reported using prescription opioids during pregnancy, among whom 21.2% reported misuse (ie, illicit use or nonmedical use of prescription opioids) [2]. Pregnant women who misuse opioids face obstetric complications, and their infants are at risk of developing neonatal abstinence syndrome (NAS), which can impede neonatal growth and increase the risk of other physical and developmental problems [3].

Study Aims

Opioid use during pregnancy is heavily stigmatized and may be legally prosecuted in many states [4,5]. Therefore, we hypothesize that women may choose to self-manage their drug use without the assistance of a health care provider, and as a result, little is known from clinical data about their health challenges and behaviors. To better understand their experiences, we examine data from an OHC, where women seek help, in part, because of the community's anonymity. This study aims to identify (1) the characteristics of women in a public OHC who self-manage their opioid use during pregnancy and (2) their self-management support needs by qualitatively analyzing personal narratives posted in the community. In doing so, we hope to identify intervention points at the policy, organizational, and technology levels that can support the unique needs of this population and achieve harm reduction.

Next, we will briefly review the related literature on the treatment of opioid misuse during pregnancy, barriers to care for pregnant women with opioid use, self-management support, and OHCs.

Treatment for Opioid Misuse During Pregnancy

Research on the best clinical practice for managing opioid use disorder (OUD) in pregnancy has significantly advanced in the past two decades [6-8]. The American College of Obstetricians and Gynecologists recommends medication-assisted treatment (MAT) over medically supervised withdrawal (because of high relapse rates associated with withdrawal), modified prenatal care elements for OUD-related health needs, and postpartum psychosocial support services [9]. Among MAT medications, buprenorphine may result in better neonatal outcomes than methadone, including higher birth weights and lower treatment times for NAS [10]. Definitions related to OUD and treatment can be found in Multimedia Appendix 1 [8-14].

Barriers to Care for Pregnant Women With Opioid Use

Survey studies and interview-based qualitative analyses have provided a contextual understanding of the barriers to care, including prenatal care and substance use treatment, encountered by pregnant women with opioid misuse. They are often of low socioeconomic status and experience significant obstacles such as lack of insurance coverage and transportation to access prenatal care, with their opioid use making it more difficult to resolve these barriers [15,16]. With regard to seeking substance use treatment in pregnancy, fear of losing child custody and concerns about being stigmatized are cited by pregnant women as barriers [17].

Self-Management Support

Health agencies in the United States and across the world have recently recognized the importance of facilitating patients' self-efficacy in managing chronic conditions. *Self-management support* focuses on empowering individuals to take active steps in managing their own chronic conditions by providing the necessary skills and confidence via interventions such as physical activities, education, and peer support [18-20]. Studies have demonstrated the effectiveness of self-management support programs, to an extent, for managing long-term conditions such as diabetes, heart failure, dementia, and Parkinson disease [21-27].

There is evidence that self-management support interventions alongside standard care can be effective for severe mental health problems [28,29], although specific interventions for substance use disorders (SUD) have not been designed or studied [30]. A search on *self-management support for substance use disorder* on PubMed yielded few relevant results at the time of writing (November 2020). Brown and Altice [31] studied themes related to self-management of MAT medications and found that online participants discussed personal experience and strategies of using unprescribed medication, distrust with health care providers, and desire to recover. Schaub et al [32] demonstrated the effectiveness of a web-based self-help intervention for participants with problematic cannabis use if it can be supplemented by brief chat counseling.

OHCs

Multiple studies point to the utility of web-based interventions in the negotiation of self-management work [27,29,33-35]. A meta-analysis found 4 mechanisms of self-management support in online groups: collective knowledge and identity building through lived experience, social support through readily accessible gifting relationships, sociability beyond illness, and online disinhibition [35]. Here, we view OHCs in their broadest sense: they are self-organizing web-based interest groups voluntarily joined by patients, caretakers, and sometimes health professionals with a shared interest in similar health conditions. OHCs can exist in dedicated forums, such as PatientsLikeMe [36] and BabyCenter [37]; in mobile health apps, such as fertility and dietary tracking apps; or on general-purpose social media platforms, such as Facebook and Reddit. They can be created and governed by health care organizations, technology companies, or patient advocacy groups.

The majority (61%) of US adults use the internet to find health information [38]. Health consumers turn to OHCs, where participants exercise collective sense making to process competing online viewpoints [39,40], to gain experiential expertise from peers that their physicians may not provide [41-44], and to formulate actionable insights for health management [44-46]. Pregnant women seek peer support from OHCs because of constrained access to health care, dissatisfaction with care received, limited offline support, and the unavailability of information from other venues [47]. In addition to information dissemination, several studies point to OHC's utility in enhancing human connection [48-51] and in helping participants overcome stigma [49,52]. Social ties formed in online groups provide space for self-management work that can improve the experience of participants with long-term illnesses [35]. As such, OHCs provide a suitable setting by which we can obtain an initial understanding of people's experiences in sensitive and stigmatized situations, capturing not only their circumstances but also their emotional states [42,45,53]. The disadvantages of OHCs, however, include their lack of quality assurance on the consumer health information provided [54], a lack of recognized credibility from health care providers [55], and possible reinforcement of negative behaviors among people in the same network [56].

Methods

Data Source

We analyzed participant-generated content from an OHC (name omitted for the protection of user privacy) that has a long-standing history and active user participation. Compared with other social media platforms, the OHC (1) is anonymous, allowing for discussion of stigmatized and sensitive health topics; (2) does not have length limits, thereby providing space for relatively detailed accounts of personal experiences; (3) has a wide range of coverage in health condition topics, including pregnancy, substance use, and pain management, so that participants are not constrained to discuss only one aspect of their health given the complex nature of gestational opioid use; and (4) has a long history that allows us to study the activities of OHC participants at the beginning of the millennium when reports of overdoses from prescribed opioids began to rise sharply [57].

Ethics and Privacy Protection

OHC content is considered public and exempt by institutional review boards as publicly available social media data per the Code of Federal Regulations (CFR) Title 45 Part 46.101 Paragraph (b) Categories of exempt human subject research (4) [58]. We took precautions in handling the OHC data because of its sensitive nature [59,60]. The data set was deidentified by removing users' screen names and assigning a randomly generated identifier number independent of the OHC. In reporting the findings, quoted sentences were removed from potentially sensitive personal information (eg, state of residence), misspellings were corrected to help mask linguistic identity, and recounts of events were paraphrased.

Data Collection

We queried the said OHC for posts made from 2000 to 2019 that contain a *pregnancy concept equivalent* and at least one OUD-related drug name. A set of pregnancy concept equivalents, namely, "pregnant, pregnancy, expecting, baby, infant, fetus, preggers, and preggy", was iteratively developed by incorporating common expressions in sample posts and synonyms to the term *pregnant*. A list of drugs was built with the generic and brand names of the top 10 prescribed opioids among commercially insured pregnant women in the United States [61] and commonly abused prescription opioids and heroin listed by the National Institute on Drug Abuse [11]. Opioid antagonists as MAT medications were also included. The list of 36 drugs can be found in Multimedia Appendix 2. Due to the nature of the string-matching query, false-positive posts were discarded during the coding process. For example, nonpregnant SUD recovery participants may use the word *expecting* in the context of anticipating an outcome.

Data Analysis

The unit of analysis is an initiating post that refers to pregnancy and opioids and does not include its comments. We performed an inductive thematic analysis on the qualifying posts following the procedures outlined by Braun and Clarke [62]. An inductive approach is driven by data without forcing emerging themes to a pre-existing coding framework [62]. The sample size was determined by the saturation principle, namely, coding was conducted until additional samples yielded no additional insights into the topic of research [63].

Two researchers (a doctoral student with a public health background and a master's student with a nursing background) iteratively annotated the same set of 200 randomly sampled posts divided into 3 coding runs. We first annotated 100 posts and recorded all appearing concepts, which were then grouped into key themes to form a codebook. For example, upon seeing many posts that described efforts of reducing the opioid dosage, we created the themes *opioid experience*, *trimester*, *recovery stage*, and *recovery method*, as it was clear that women usually mentioned their opioid experience and stage of gestation as a context for discussing the recovery methods and their recovery progress. Second, we annotated 50 additional posts and refined the codebook by placing similar concepts together. For example, inquiries on neonatal withdrawal were combined with general questions about the drug safety of opioids as *adverse effects of gestational opioid use*. Third, we annotated another 50 posts to confirm that no new concepts emerged from the annotation to reach saturation. The development of the codebook was supervised by an experienced researcher in human-computer interaction. The definitions and exemplary quotes of the key themes and concepts are provided in Table 1. Among the variables, up to 3 (the maximum number found in the posts) emotions and self-management support needs were annotated, as participants can express more than one concept (emotion or concern) in the same post. To measure interrater reliability, the annotations have a Cohen kappa of 0.863, which suggests a high level of agreement. Interpretative differences were discussed among the 3 researchers and resolved.

Table 1. Codebook.

Study aim, theme, and concept definition	Example
Study aim 1	
Opioid experience	
Opioid naïve: • Temporary use of opioid prescriptions for acute pain lasting fewer than 3 months	• "I had stomach pains and went to the emergency room. The doctor gave me morphine." [P3660]
Opioid misuse: • Meeting one or more DSM-V[a] diagnostic criteria for OUD[b] [64] • Receiving treatment for OUD	• "I have been self-medicating myself a total of 18.75 mg daily for about a year and a half. When I stopped, I experienced withdrawal symptoms! Pretty intense ones too." [P90] • "I'm currently on methadone for an opiate addiction." [P1600]
Unable to determine	• "I am 3 weeks pregnant and on Norco, Flexeril, Xanax. I have been taking them for 2 years." [P1830]
Recovery stage	
Precontemplation: • No mention of interest in reducing the opioid dosage	• "I've been taking Vicodin and Percocet and I'm 20 weeks pregnant." [P5357]
Contemplation: • Expressing interest in reducing the dosage	• "I am ready and would like to quit the suboxone cold turkey." [P4191]
Action: • Describing experience during withdrawal or relapse	• "I am on day 2 of detoxification. How long till I feel better?" [P4074]
No misuse	• "I am 38 weeks pregnant and my Ob prescribed me Percocet for kidney stones. I passed the stone today, so I won't be needing the pain killers anymore." [P6315]
Recovery method	
Tapered withdrawal (self-managed): • Describing preference for gradually reducing the dosage of opioids	• "I have been detoxing for 2 months and have gone down to 29 mgs." [P1649]
Sudden discontinuation (self-managed): • Using expressions that indicate full discontinuation of opioids	• "I have 2 weeks left till my delivery date I stopped taking the Norco today." [P1339]
Undecided self-recovery: • Not specifying a particular method but expressing interest in dosage reduction on one's own	• "I know I can't do this cold turkey and am unsure of my will power to taper." [P6576]
Professional treatment: • Receiving substance use treatment from professionals	• "I take the methadone daily at a clinic near where I live." [P1600]
MAT[c] (source unknown): • Using MAT medications from unspecified sources	• "Is it safe to use Suboxone while being pregnant?" [P6066]
Not applicable: • Women who did not have opioid misuse or not pursuing recovery	• "I'm 27 weeks pregnant and have peed blood, and my sides where my kidneys are have been hurting really bad. I'm trying not to take the Tylenol with codeine, but I might have to if it keeps getting worse." [P17]
Trimester	
First trimester: • 0-13 weeks	• "I have just found out I am 6 weeks pregnant." [P6434]
Second trimester: • 14-27 weeks	• "I am currently 22 weeks pregnant with my second child." [P3644]

Study aim, theme, and concept definition	Example
Third trimester: • 28+ weeks	• "I am 39 weeks pregnant." [P5896]
Unspecified	• "I think I'm on the right path now, but I'm scared it will still be in my system when I have the baby." [P460]
Study aim 2	
Self-management support needs—informational	
Potential adverse effects of gestational opioid use • Neonatal withdrawal • Open-ended inquiry	• "I was just wondering if my baby could possibly withdrawal from these?" [P6342] • "Has anyone taken this and (was) baby ok?" [P217]
Self-managed withdrawal	• "Has anyone had any good stories with getting lowered slowly off it?" [P12765]
Pain management safe for pregnancy	• "I am prescribed Percocet for pain. Is it bad for baby, is there a best alternative?" [P1283]
Legal procedures	• "I need to know if they can take my baby because of this prescription showing up?" [P5143]
Navigating offline support systems: • Looking for recommendations of treatment facilities • Interacting with providers • Interacting with caretakers	• "If anyone knows of an OBGYN that takes [insurance name] and is familiar with my situation, or a pain management doc that deals with pregnancies and also takes my insurance...just SOMEWHERE to start." [P141] • "I'm not sure what to do anymore. I don't want my doc to think I'm abusing them or selling then because I swear I'm not!" [P8] • "I just don't want to bring up my drug use to my mom because I know it's going to hurt her. I don't know what to do or what to say that doesn't have my mom worry about my baby's healthy." [P3413]
Other pregnancy concerns	• "I am not sure if I have felt the baby move yet, I am 18 weeks today. Is that ok?" [P6689]
Appropriateness of tapering schedule	
Appropriate: • The tapering plan is consistent with clinical guidelines	• No instances were found.
Inappropriate: • The tapering plan is too rapid	• "I took 3, 30 mg for the past few days and I took 2, 30 mg today. I am going to take 1, 30 mg tomorrow, 15 the next and half that the following." [P6844]
Unclear: • Not enough information to determine appropriateness	• "Can I ween myself off of Methadone slowly, very slowly?" [P3391]
Self-management support needs—emotional	
Seeking emotional support	• "Some days I have a hard time staying positive. Please if anyone is available to talk with me, I would really appreciate it." [P3899]
Sentiment	

Study aim, theme, and concept definition	Example
Negative sentiments: • Fear • Shame • Anxiety • Despair	• "I am terrified of losing him." [P3545] • "I am very ashamed." [P90] • "I am planning to cut down and then quit because I'm worried, even though doctors say they are ok." [P369] • "I am so ashamed, scared, and lonely! I feel hopeless!" [P4608]
Mixed sentiment: • Cautious optimistic	• "I feel much more optimistic about beating my addiction for good after coming to this forum. It's hard when you have no one to talk to or share experiences with." [P2137]
Positive sentiment: • Positive emotions, such as hope and love	• No instances were found.

[a]DSM-V: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

[b]OUD: opioid use disorder.

[c]MAT: medication-assisted treatment.

Furthermore, women's common beliefs were compared with the scientific literature and clinical guidelines to discern any divergence. An experienced clinical psychologist annotated whether misconception existed compared with clinical guidelines in the self-guided withdrawal plans described by participants [65].

Data Reporting

Themes from inductive coding were grouped in the results to address the 2 study aims. First, the characteristics of OHC participants (study aim 1) are represented by themes coded under *Opioid experience*, *Recovery stage*, *Recovery method*, and *Trimester*. Second, the self-management support needs (study aim 2) are represented by concepts coded under self-management support *needs—informational* and self-management support *needs—emotional* and complemented by the themes *Appropriateness of tapering schedule* and *Sentiment*.

Results

Metadata

The search query yielded 3559 posts between 2000 and 2019. Posts appeared in 201 subgroups related to substance misuse (2096/3559, 58.89%), pregnancy (680/3559, 19.11%), and others (783/3559, 22.00%), such as neurology and back pain. The mean number of drug names mentioned per post was 1.3 (SD 0.7). The mean character count per post was 1411.2 (SD 1313.7).

Study Aim 1: Characteristics of Pregnant Women With Opioid Use in OHC

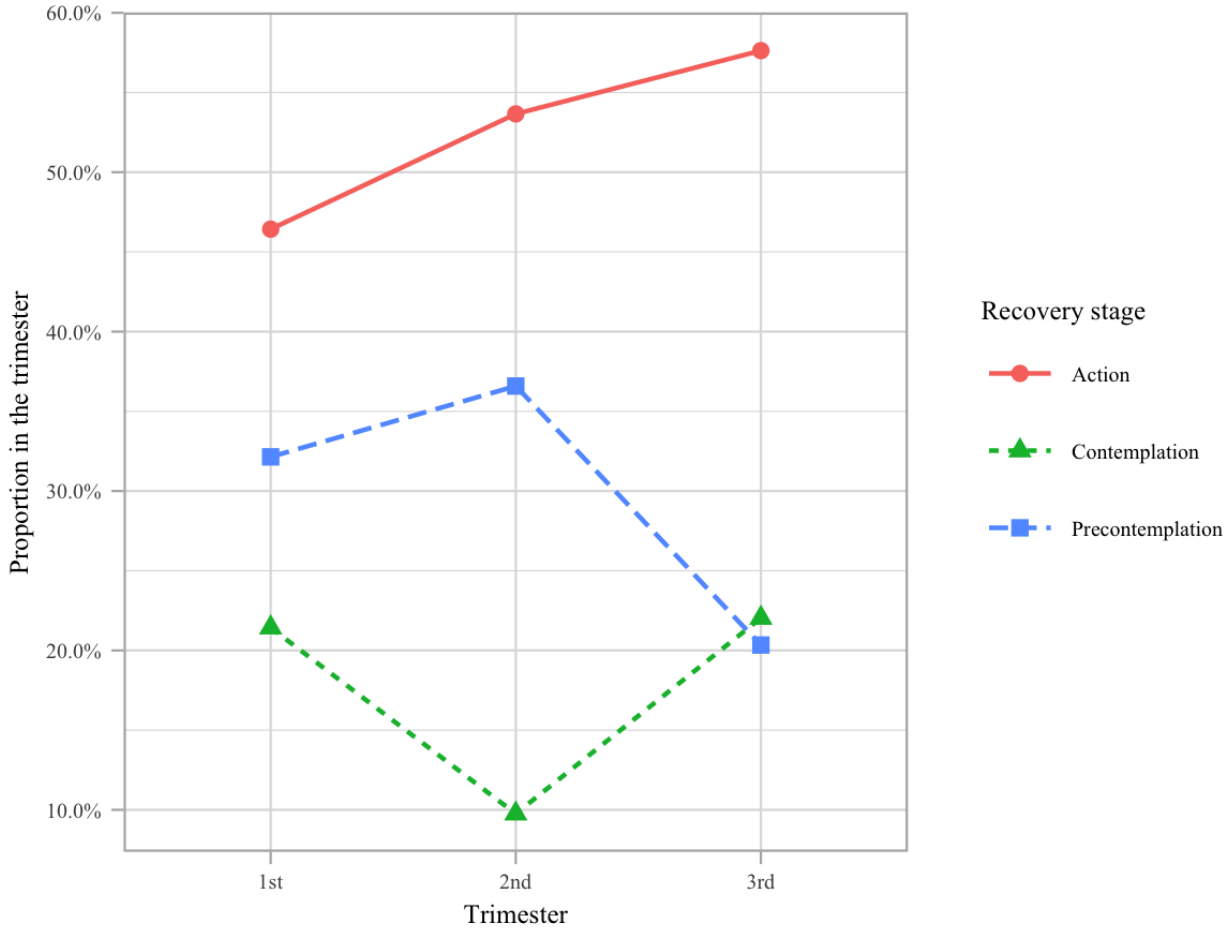
The majority (150/200, 75.0%) of women who took opioids during pregnancy in the OHC (*the study population*) met one or more criteria from the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [64] for a potential OUD diagnosis, which we refer to as *opioid misuse* as we were unable to make a formal clinical diagnosis. Close to half (94/200, 47.0%) of the study population were in the process of pursuing dosage reduction (ie, action stage), with another 31 participants (31/200, 15.5%) considering but not yet initiating a reduction in dosage (ie, contemplation stage; Table 2). In terms of recovery method, self-managed withdrawal was more common than professional treatment (P<.001). This indicates that women in the OHC primarily elect to self-manage their attempts at dosage reduction during pregnancy.

Gestationally, the women were primarily in their first or third trimesters, and notably, the percentage of women not pursuing recovery (ie, precontemplation stage) decreased (P=.11) and the percentage of those in action increased (P=.16) from the first to the third trimester, although the difference was not statistically significant (Figure 1).

Table 2. Characteristics of women inquiring about opioid use in the online health communities (N=200).

Characteristic	Participants, n (%)
Opioid experience	
Opioid naïve	22 (11.0)
Opioid misuse	150 (75.0)
Unable to determine	28 (14.0)
Recovery stage	
Precontemplation	51 (25.5)
Contemplation	31 (15.5)
Action	94 (47.0)
No misuse	24 (12.0)
Recovery method	
Self-managed withdrawal	
Tapered withdrawal	40 (20.0)
Sudden discontinuation	35 (17.5)
Undecided	15 (7.5)
Professional treatment	29 (14.5)
Medication-assisted treatment (unknown sources)	6 (3.0)
Not applicable	75 (37.5)
Trimester	
First	61 (30.5)
Second	49 (24.5)
Third	66 (33.0)
Unspecified	24 (12.0)

Figure 1. Proportions of the recovery stages by trimester.



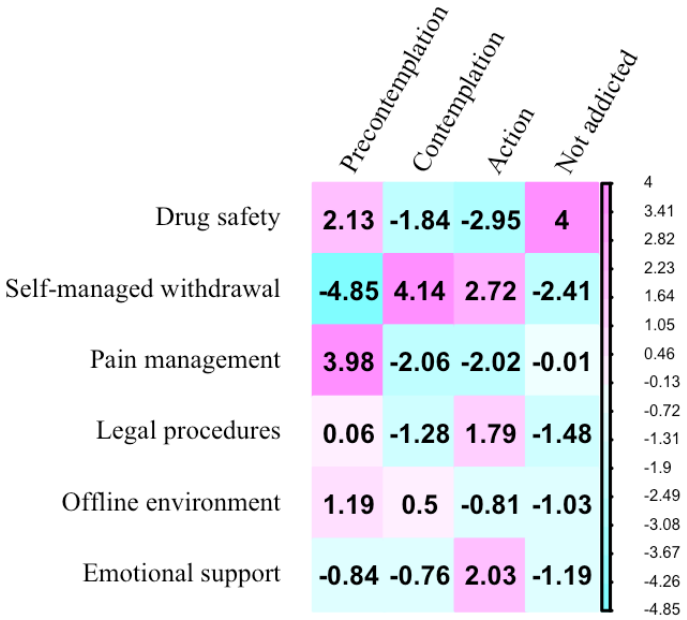
Study Aim 2: Self-Management Support Needs

Following thematic analysis of the web posts, 6 primary themes of self-management support needs were identified, including information needs for understanding (1) the potential adverse effects of gestational opioid use, (2) self-led withdrawal, (3) the safety of continued opioid use for pain management during pregnancy, (4) legal procedures related to child protection, (5) navigating offline health care systems, and (6) needs for emotional support (Table 3). Other pregnancy concerns and posts with unspecified support needs were excluded from further analysis because they do not pertain to opioid use. The women had different self-management support needs according to their recovery stages (χ^2_{15}=69.5; P<.001). The absolute value of a standardized residual (r_{std}) greater than 2 indicates a strong (dis)association (Figure 2), on which we will elaborate next.

Table 3. Self-management support needs expressed in the online health community postings (N=200).

Themes and concepts	Participants, n (%)
Self-management support needs–informational	
Potential adverse effects of gestational opioid use	99 (49.5)
Self-managed withdrawal	70 (35.0)
Pain management safety during pregnancy	20 (10.0)
Legal procedures	18 (9.0)
Navigating offline support systems	9 (4.5)
Self-management support needs–emotional	
Seeking emotional support	12 (6.0)
Excluded	
Other pregnancy concerns	5 (2.5)
Unspecified	4 (2.0)

Figure 2. Standardized residuals of associations between self-management support needs and recovery stages.



Potential Adverse Effects of Gestational Opioid Use

The most common concern (99/200, 49.5%) of the study population was the potential adverse effects of opioids on the fetuses. In particular, the questions on drug safety have a high level of homogeneity and can be simply divided into 2 categories: general inquiry as in "was your baby okay?" [P3614] and specific inquiry on the potential of neonatal withdrawal as in "I am scared to death my baby will have withdrawals" [P16]. These concerns were primarily seen among those who did not consider treatment for opioid misuse (r_{std}=2.13) or did not have a misuse (r_{std}=4.00), whereas those in active pursuit of recovery were least concerned with the drug safety in their posts (r_{std}=−2.95; Figure 2).

Self-Managed Withdrawal

Questions about how to reduce opioid dosage during pregnancy were the second most common (70/200, 35.0%), primarily among those contemplating (r_{std}=4.14) or undergoing (r_{std}=2.72) treatment of opioid misuse (Figure 2). These concerns included (1) comparing strategies to reduce opioid dosage, (2) discussing withdrawal schedules, (3) enduring withdrawal-related hardships, and (4) dealing with the aftermath of relapses.

First, women contemplating dosage reduction used the OHC as a sounding board to plan their course of action in the absence of professional advice. They debated the risks and benefits of withdrawal during pregnancy and how to do so safely. They were concerned about the impact on the fetus if withdrawal symptoms occurred during their pregnancy:

> *I'm scared about the withdrawals of MS Contin, for which I'm fully prepared, but scared it will harm my child?* [P6419]

They also compared sudden discontinuation to tapered withdrawal:

> *I considered quitting cold turkey, but I have read that it is a bad idea. Would tapering be a good idea? What is a good taper method?* [P2325]

Although this shows some women were aware of the disadvantages of sudden discontinuation, tapered withdrawal (40/200, 20.0%) and sudden discontinuation (35/200, 17.5%) were equally popular among the study population (*P*=.61).

Second, once committed to decreasing their opioid dosage, the women may describe a detailed tapering schedule in the OHC to solicit peer feedback:

> *Should I just do half of what I have been for a few days, then cut it in half again, do that for a couple days, then go for nothing?! I really don't know what to do, that's why I am asking for help.* [P1628]

Of all the women who indicated an interest in tapered withdrawal, only 23% (9/40) provided specific information on how they intended to do so, suggesting that most did not have a structured plan on how to taper. Furthermore, *all* of these plans were deemed too rapid (Table 4) based on clinical guidelines and could increase the risk of painful withdrawal symptoms [65].

Table 4. Appropriateness of tapering schedule (N=40).

Appropriateness	Participants, n (%)
Unclear	31 (78)
Inappropriate (too rapid)	9 (22)
Appropriate	0 (0)

Third, after initiating withdrawal, the women turned to the OHC for practical advice and emotional support when undergoing the hardship of withdrawal, sometimes stemming from the lack of a proper tapering schedule:

> *Just started detox on tramadol today. I feel like hell and my legs are killing me! I'm 26 hours into it. How much more do I need to endure?* [P2423]

As a result of significant withdrawal symptoms, they were prone to setbacks and relapses:

> *Unfortunately, I relapsed and slowly started using until now (37 weeks). I have stopped a couple of times but have had issues because of withdrawal fears.* [P5992]

To combat the demoralizing impact of relapses, some women looked to the OHC to hold them accountable:

> *I am just looking for some support, accountability and encouragement as I feel so alone, scared and terrible about my relapse and lying to my love.* [P3545]

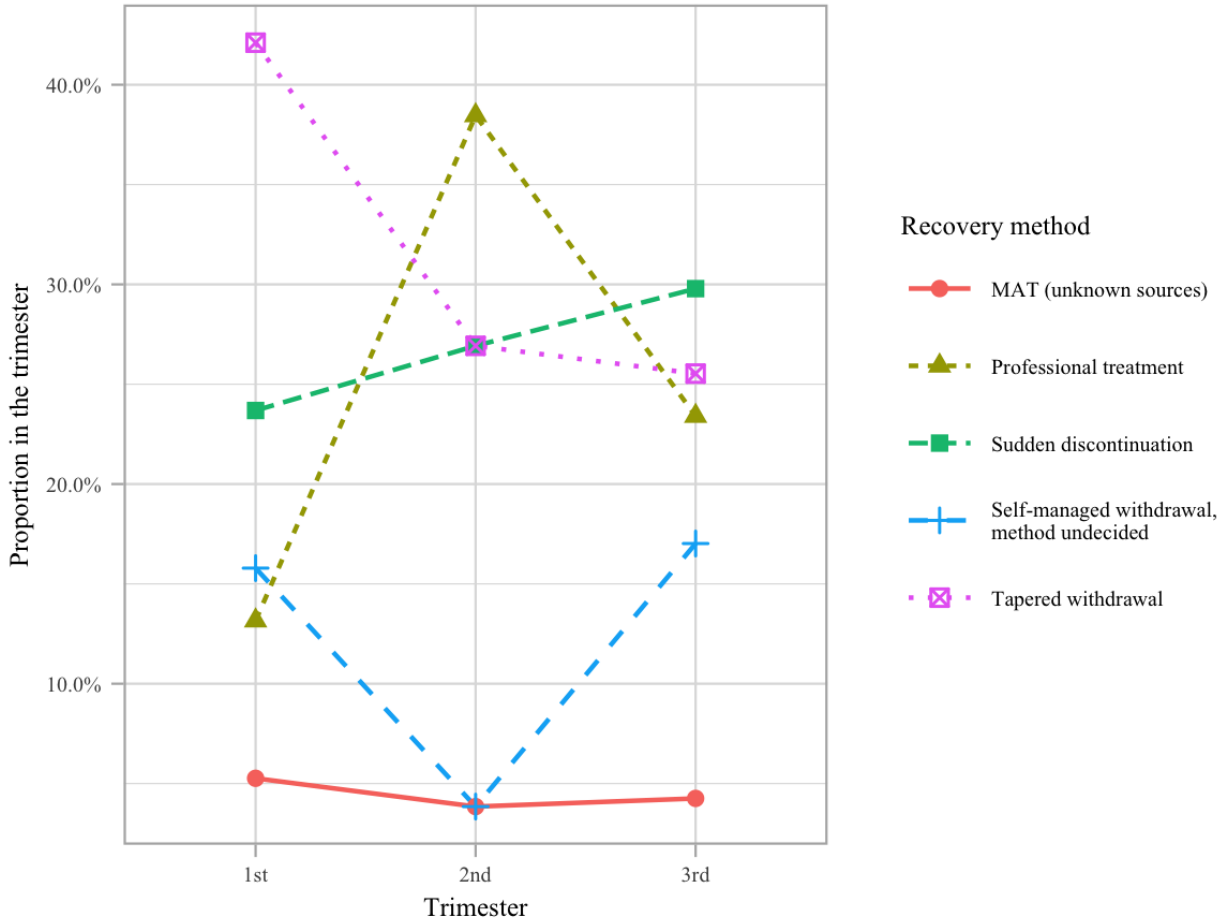
These experiences highlight the absence of structure and support with self-managed withdrawal as compared with being in a clinical program where physician supervision and evidence-based therapies are provided to help patients manage their dosage reduction process more effectively.

Fourth, the repeated attempts and failures to decrease opioid use during pregnancy were not only physically taxing and emotionally draining but also meant that the women with clandestine opioid use would have no other perceived choice but to resort to sudden discontinuation when the date of delivery approached in hope of a negative drug test, despite the dangers of doing so abruptly. Women in this position asked:

> *How long will it take for the test to be negative if I stop today?* [P5891]

The commonness of this risky approach is evident in the increasing percentage (P=.70) of women attempting sudden discontinuation of dosage and the decreasing percentage (P=.17) of tapered withdrawal during the third trimester (Figure 3), although the changes were not statistically significant.

Figure 3. Proportions of the recovery methods by trimester.



Pain Management

Women using opioids for chronic pain management before becoming pregnant faced the dilemma of leaving their severe pain untreated or risking side effects to their fetus. They had difficulty transitioning to gestational opioid use:

> *I have been in pain management for 3 years due to a dislocated hip. Apparently neither my OB nor primary care doctor have ever dealt with "my situation" before.* [P141]

In looking for safe pain medications to take during pregnancy, they (20/200, 10.0%) viewed peers' experience as empirical evidence and second opinions to professional advice:

> *I have 4 discs missing in my spine. I am now 4 months pregnant and my doctor has taken me off everything!!! Does anyone know what is safe for the baby?* [P6486]

As they continue to take opioids during pregnancy, concerns related to pain management are mostly associated (r_{std}=3.98) with those in the precontemplation stage (Figure 2).

Understanding Legal Procedures

Concerns related to Child Protective Services (CPSs) procedures were present in 9.0% (18/200) of the posts. These concerns were more common among women in the recovery stage of action (r_{std}=1.79) than in other stages (Figure 2). Although the regulation and laws related to child protection from drug use are obscure to laypersons, the vague idea of losing parental guardianship looms large. Pregnant women with opioid misuse in the OHC described intense fear of being identified as *using* and consequently losing parental guardianship of their newborn and the goodwill of their direct support systems:

> *I am scared to death of having to deal with child protection service especially since none in my family knows about any of this and the fear of my child being taken away from me.* [P16]

They inquired about the role of hospitals in reporting drug use instead of directly asking their providers:

> *Will child services still get involved? I have heard that doctors do these types of things and contact these types of people behind your back?* [P3776]

Notably, those who used unprescribed MAT medications to self-treat OUD (6/200, 3.0%) may still be held accountable for illegal possession and opioid use during pregnancy.

Navigating Offline Support Systems

A few women (9/200, 4.5%) also requested strategies for navigating their offline support systems, including recommendations for treatment facilities (4/200, 2.0%), advice for interacting with health care providers (4/200, 2.0%), and disclosing opioid use with their family members (1/200, 0.5%). These concerns were not particularly associated with one recovery stage. Accessing specialized prenatal care for women with opioid use or finding substance treatment programs that accept pregnant women presented a challenge. Locating a specialty program proved to be difficult even with an obstetrician's referral:

> *My obstetrician had made me an appointment with a therapist, who tells me she doesn't even see pregnant women! I got some numbers for the methadone clinic and other places but yet again no one will help a pregnant woman.* [P3855]

This post was made after 2010, which is a decade after the passing of the Drug Addiction Treatment Act (DATA) of 2000, which allows trained physicians including obstetricians to treat opioid dependency with MAT medications. Women also consulted the OHC for strategies on how to best interact with their providers and caretakers to explain personal circumstances involving opioid usage and building rapport (examples are given in Table 1).

Seeking Emotional Support

Besides the needs for informational support, a few women (12/200, 6.0%) explicitly requested emotional support:

> *Please send me some words of encouragement.* [P3545]

In contrast, a high number of posts (117/200, 58.5%) described negative emotions, including fear, shame, anxiety, and despair (Table 5; examples are given in Table 1). In other words, only about 1 out of 10 women who experienced negative emotions sought to address their emotional needs with the help of the OHC. Importantly, expressing negative emotions did not modify the women's likelihood of pursuing (ie, action or contemplation) dosage reduction (χ^2_1=0.1; *P*=.75).

Table 5. Sentiment (N=200).

Sentiment	Participants, n (%)
Negative	117 (58.5)
Not specified	82 (41.0)
Mixed	1 (0.5)
Positive	0 (0.0)

Discussion

Principal Findings

The majority of pregnant women in the OHC exhibited signs of opioid misuse, with approximately two-thirds of them pursuing recovery. Self-managed withdrawal of opioid use was more common than professional treatment. The following 6 identified themes highlighted women's self-management support needs: (1) providing clarity on the impact of opioid drugs on pregnancy; (2) providing clinically validated information on how to scientifically reduce opioid dosage; (3) providing guidelines on safe pain management practice during pregnancy; (4) providing information on local CPSs procedures, including the hospital's role in reporting; (5) providing strategies for interacting with and obtaining support from offline support systems; and (6) providing emotional support for those experiencing negative emotions.

The study population relied heavily on the OHC to provide guidance in the absence of professional care and in-person support, which differentiates them from other patient groups

that usually use online support groups as a supplement to traditional health care services or in-person support groups [66,67]. Formal evaluation and proper treatment recommendations were lacking for pregnant women in the OHC who chose self-managed withdrawal. Despite their resolve to reduce opioid dosage, women were vulnerable to the pitfalls of misinformation (eg, overly aggressive tapering schedules) and could easily experience relapses that may cause distress in both the mother and fetus.

Indirect Emotional Support

Experiencing negative emotions is commonplace for the study population, but explicitly requesting emotional support is not. Furthermore, although positive emotions are shown to be a facilitator of self-efficacy, which is a key construct in the social cognitive theory for effecting health behavior change [68,69], the opposite may not be true: experiencing negative emotions does not modify the women's likelihood of pursuing dosage reduction. This contrast, on the one hand, shows that women with opioid use or misuse during pregnancy were preoccupied with seeking information to resolve their predicament, and on the other hand, may suggest that their emotional needs were met by the OHC in indirect ways. First, information seeking is frequently used as a coping response because it helps assess the degree of threat associated with a stressor, thereby reducing uncertainty in health care [69,70]. Second, although not directly soliciting emotional support, participants in the OHC voluntarily shared detailed accounts of their offline experiences while seeking information on self-managed withdrawal, pain management, and strategies for navigating health care environments. Emotional regulation in the form of venting feelings is within the underpinning of the transactional model of stress and coping [70,71]. Cathartic release and negation of offline frustration are also themes related to the negotiation of self-management support [35].

Challenging the Scope of Self-Management Support

It is worth noting that self-management support interventions are typically developed by health care professionals to complement standard care. For example, self-management support for mental health typically focuses on patient education, medication adherence, relapse prevention, and coping strategies [28]. Among the 6 self-management support needs identified, the second need for patient education on how to scientifically reduce opioid dosage may challenge the realm of what is commonly accepted for self-management support in that self-managed withdrawal implies evading standard care, instead of complementing it. Self-managed withdrawal without professional supervision can be dangerous and should be discouraged. Here, we take the view that until there are enough specialized resources to treat pregnant women with opioid misuse and done so without legal penalization, there will always be women compelled to pursue self-managed withdrawal. Therefore, harm reduction via patient education as to how to safely taper is imperative.

Policy Considerations

Although DATA was passed in 2000, women in the OHC still reported difficulty in obtaining MAT from their obstetricians' office years later. To facilitate the initiation of recovery for pregnant women with opioid misuse, it is imperative to increase the number of obstetricians who are waivered buprenorphine prescribers and to increase the number of opioid treatment programs that cater to pregnant women. As of January 2020, approximately 10% of US physicians have received the DATA waiver [72]. Coverage in rural areas is particularly needed. For example, only 53% of outpatient buprenorphine prescribers accepted pregnant patients in Appalachian states as of 2017 [73]. Moreover, only 24% of opioid treatment facilities offer special programs for pregnant or postpartum women [1].

The popularity of self-management in the OHC highlights women's needs for support in reducing their opioid dosage and their fear of seeking professional care. This indicates that legal penalization can be detrimental to the well-being of both the mother and child, as women avoid prenatal visits [4]. Many pregnant women presume that they will face negative consequences if they disclose drug use to their obstetricians, missing an important window in which to initiate recovery. Universal screening of substance use for pregnant women without legal implications may help dissolve the distrust between some patients and their providers.

Technology Design for Harm Reduction

Although technology may not be able to directly change the medical and legal landscape, it can be used to tackle challenges faced by pregnant women with opioid misuse. Digital interventions have demonstrated small to modest effects in supporting people in recovery from SUDs [74,75]; however, only a minority of evidence-based self-help interventions have functional websites for general use [76]. A study on popular alcohol reduction apps found that common behavior change techniques employed facilitate self-recording, provide information on the consequences of excessive alcohol use, and provide feedback on performance [77]. As mentioned earlier, the social cognitive theory offers a framework to create positive behavior changes [68,69]. *Facilitation* and *self-regulation* are 2 concepts that are particularly applicable to better technological designs in the context of this study.

Facilitation refers to providing tools and resources to make new behaviors easier to establish. OHCs can facilitate (1) geo-specific information dissemination and (2) clinically validated tapering schedules for those who opt for self-management. Specifically, given the regional variations in law enforcement and resources of specialized care, organizing information specific to geographic areas may facilitate online discussions relevant to participants' local environments. For those opting for self-managed withdrawals, we envision an online calculator that can account for women's historical opioid dosage and gestational stage and generate a personalized tapering schedule based on current guidelines (albeit while expressing strong encouragement to bring one's opioid use to the attention of their health care professionals).

Furthermore, self-regulation refers to controlling oneself through self-monitoring, goal setting, feedback, and the enlistment of social support. Although undergoing withdrawal, women in the OHC often lack structured social support that can hold them accountable and support them in their efforts to stay on track.

Peer-led 12-step groups have been shown to improve accountability and recovery prospects for participants [78]. We envision OHC's incorporating the structured aspect of effective peer support programs by creating an environment in which participants can perform daily check-ins, display badges of withdrawal progress, and easily reach out to peers for support and accountability.

Limitations and Future Work

The findings of this study should be interpreted with limitations. First, the reported self-recovery trends are representative of those seeking help in the OHC. The percentage of self-managed withdrawal in the general population may be lower than that reported in this study, as those opting for self-management may have a greater propensity for participating in the OHC discussions to seek help. Our findings, however, are meaningful in better understanding the OHC population that appears to require relevant, clinically validated information. Second, only 200 posts for the OHC were analyzed. Thematic analyses are commonly applied to sociobehavioral studies using semistructured interviews with an average of 30 participants (SD 18.7) [79]. Qualitative studies of online content have varying sample sizes, usually ranging from 100 to 2000 [31,41,43,45,46,50,80]. We iteratively coded 200 posts and reached concept saturation within this number. In other words, had new concepts emerged during the annotation of the last 50 randomly sampled posts, we would continue sampling additional posts. Third, comments in response to the initiating posts by women with opioid misuse were not analyzed, as this study focused on the beliefs and actions of those attempting self-management at the moment of posting. Previous research shows that most comments in OHCs for chronic health conditions have an element of social support, primarily including validation and empathy [80]. In light of our observation on the imbalance of requests for informational and emotional support, future research should examine how the OHC audience responds to the posts. Fourth, qualitative analysis was performed on a cross-sectional rather than a longitudinal sample set. A future direction may be to follow OHC participants throughout the course of their pregnancy and postpartum to better understand the outcome of their proposed self-management work.

Conclusions

OHCs provide vital self-management support for pregnant women with opioid use or misuse. Women pursuing self-managed dosage reduction are prone to misinformation and repeated relapses, which can result in extreme measures to avoid testing positive for drug use at labor. The study findings provide evidence for public policy considerations, including universal screening of substance use for pregnant women, emphasis on treatment rather than legal punishment, and further expansion of the DATA waiver training program. The improvement of online platforms that can organize geo-relevant information, dispense clinically validated withdrawal schedules, and offer structured peer support is envisioned for harm reduction among pregnant women who opt for self-management of opioid misuse.

Acknowledgments

This work was supported in part by the National Science Foundation under Grant NSF-1741306, Grant IIS-1650531, and Grant DIBBS-1443019. Any opinions and conclusions or recommendations expressed in this study are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions related to opioid use and treatment.
[DOCX File , 22 KB - jmir_v23i2e18296_app1.docx]

Multimedia Appendix 2
List of 36 drug names related to opioid use.
[DOCX File , 14 KB - jmir_v23i2e18296_app2.docx]

References

1. Smith K, Lipari R. Women of childbearing age and opioids. The CBHSQ Report. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2017. URL: https://www.samhsa.gov/data/sites/default/files/report_2724/ShortReport-2724.pdf [accessed 2021-01-05]
2. Ko JY, D'Angelo DV, Haight SC, Morrow B, Cox S, Salvesen von Essen B, et al. Vital Signs: Prescription Opioid Pain Reliever Use During Pregnancy - 34 U.S. Jurisdictions, 2019. MMWR Morb Mortal Wkly Rep 2020 Jul 17;69(28):897-903 [FREE Full text] [doi: 10.15585/mmwr.mm6928a1] [Medline: 32673301]
3. Yeoh SL, Eastwood J, Wright IM, Morton R, Melhuish E, Ward M, et al. Cognitive and Motor Outcomes of Children With Prenatal Opioid Exposure: A Systematic Review and Meta-analysis. JAMA Netw Open 2019 Jul 03;2(7):e197025 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.7025] [Medline: 31298718]

4. Krans EE, Patrick SW. Opioid Use Disorder in Pregnancy. Obstetrics & Gynecology 2016;128(1):4-10. [doi: 10.1097/aog.0000000000001446]
5. Livingston JD, Milne T, Fang ML, Amari E. The effectiveness of interventions for reducing stigma related to substance use disorders: a systematic review. Addiction 2012 Jan;107(1):39-50 [FREE Full text] [doi: 10.1111/j.1360-0443.2011.03601.x] [Medline: 21815959]
6. Klaman S, Isaacs K, Leopold A, Perpich J, Hayashi S, Vender J, et al. Treating Women Who Are Pregnant and Parenting for Opioid Use Disorder and the Concurrent Care of Their Infants and Children: Literature Review to Support National Guidance. J Addict Med 2017;11(3):178-190 [FREE Full text] [doi: 10.1097/ADM.0000000000000308] [Medline: 28406856]
7. Winklbaur B, Kopf N, Ebner N, Jung E, Thau K, Fischer G. Treating pregnant women dependent on opioids is not the same as treating pregnancy and opioid dependence: a knowledge synthesis for better treatment for women and neonates. Addiction 2008 Sep;103(9):1429-1440. [doi: 10.1111/j.1360-0443.2008.02283.x] [Medline: 18783498]
8. Stover MW, Davis JM. Opioids in pregnancy and neonatal abstinence syndrome. Semin Perinatol 2015 Dec;39(7):561-565 [FREE Full text] [doi: 10.1053/j.semperi.2015.08.013] [Medline: 26452318]
9. Committee on Obstetric Practice. Committee Opinion No. 711: Opioid Use and Opioid Use Disorder in Pregnancy. Obstet Gynecol 2017 Aug;130(2):e81-e94. [doi: 10.1097/AOG.0000000000002235] [Medline: 28742676]
10. Laslo J, Brunner J, Burns D, Butler E, Cunningham A, Killpack R, et al. An overview of available drugs for management of opioid abuse during pregnancy. Matern Health Neonatol Perinatol 2017 Feb 10;3(1):4 [FREE Full text] [doi: 10.1186/s40748-017-0044-2] [Medline: 28203387]
11. Commonly Abused Drugs Charts. National Institute on Drug Abuse. 2019. URL: https://www.drugabuse.gov/drugs-abuse/commonly-abused-drugs-charts [accessed 2021-01-05]
12. Module 5: Assessing and Addressing Opioid Use Disorder (OUD). Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/drugoverdose/training/oud/accessible/index.html [accessed 2021-01-20]
13. Worley J. Recovery in substance use disorders: what to know to inform practice. Issues Ment Health Nurs 2017 Jan;38(1):80-91. [doi: 10.1080/01612840.2016.1245375] [Medline: 27901625]
14. Propublica. How States Handle Drug Use During Pregnancy. URL: https://projects.propublica.org/graphics/maternity-drug-policies-by-state [accessed 2021-01-20]
15. Metz VE, Brown QL, Martins SS, Palamar JJ. Characteristics of drug use among pregnant women in the United States: Opioid and non-opioid illegal drug use. Drug Alcohol Depend 2018 Feb 01;183:261-266 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.11.010] [Medline: 29310077]
16. Roberts SCM, Pies C. Complex calculations: how drug use during pregnancy becomes a barrier to prenatal care. Matern Child Health J 2011 Apr 16;15(3):333-341 [FREE Full text] [doi: 10.1007/s10995-010-0594-7] [Medline: 20232126]
17. Frazer Z, McConnell K, Jansson LM. Treatment for substance use disorders in pregnant women: Motivators and barriers. Drug Alcohol Depend 2019 Dec 01;205:107652. [doi: 10.1016/j.drugalcdep.2019.107652] [Medline: 31704383]
18. Self-Management Support. Agency for Healthcare Research Quality. 2020. URL: http://www.ahrq.gov/ncepcr/tools/self-mgmt/self.html [accessed 2021-01-05]
19. Support and Education for Patient Disease Management. Cent Dis Control Prev. URL: https://www.cdc.gov/dhdsp/pubs/guides/best-practices/self-management.htm [accessed 2021-01-05]
20. Module 2: Self-Management Support (SMS). U.S. Department of Health & Human Services. URL: https://www.hhs.gov/ash/about-ash/multiple-chronic-conditions/education-and-training/curriculum/module-2-self-management-support/index.html [accessed 2021-01-05]
21. Allegrante JP, Wells MT, Peterson JC. Interventions to Support Behavioral Self-Management of Chronic Diseases. Annu Rev Public Health 2019 Apr 01;40:127-146. [doi: 10.1146/annurev-publhealth-040218-044008] [Medline: 30601717]
22. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth Interventions to Support Self-Management of Long-Term Conditions: A Systematic Metareview of Diabetes, Heart Failure, Asthma, Chronic Obstructive Pulmonary Disease, and Cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
23. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027. [doi: 10.1177/1932296817713506] [Medline: 28560898]
24. Kessler D, Liddy C. Self-management support programs for persons with Parkinson's disease: An integrative review. Patient Educ Couns 2017 Oct;100(10):1787-1795. [doi: 10.1016/j.pec.2017.04.011] [Medline: 28465112]
25. Galdas P, Fell J, Bower P, Kidd L, Blickem C, McPherson K, et al. The effectiveness of self-management support interventions for men with long-term conditions: a systematic review and meta-analysis. BMJ Open 2015 Mar 20;5(3):e006620-e006620 [FREE Full text] [doi: 10.1136/bmjopen-2014-006620] [Medline: 25795688]
26. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]

27. Huis In Het Veld JG, Willemse BM, van Asch IF, Groot Zwaaftink RB, Verkade P, Twisk JW, et al. Online Self-Management Support for Family Caregivers Dealing With Behavior Changes in Relatives With Dementia (Part 2): Randomized Controlled Trial. J Med Internet Res 2020 Feb 25;22(2):e13001 [FREE Full text] [doi: 10.2196/13001] [Medline: 32130142]
28. Lean M, Fornells-Ambrojo M, Milton A, Lloyd-Evans B, Harrison-Stewart B, Yesufu-Udechuku A, et al. Self-management interventions for people with severe mental illness: systematic review and meta-analysis. Br J Psychiatry 2019 May;214(5):260-268 [FREE Full text] [doi: 10.1192/bjp.2019.54] [Medline: 30898177]
29. Poole R, Smith D, Simpson S. How Patients Contribute to an Online Psychoeducation Forum for Bipolar Disorder: A Virtual Participant Observation Study. JMIR Ment Health 2015 Jul;2(3):e21 [FREE Full text] [doi: 10.2196/mental.4123] [Medline: 26543925]
30. Lucock M, Gillard S, Adams K, Simons L, White R, Edwards C. Self-care in mental health services: a narrative review. Health Soc Care Community 2011 Nov;19(6):602-616. [doi: 10.1111/j.1365-2524.2011.01014.x] [Medline: 21749527]
31. Brown S, Altice FL. Self-management of buprenorphine/naloxone among online discussion board users. Subst Use Misuse 2014 Jun;49(8):1017-1024. [doi: 10.3109/10826084.2014.888449] [Medline: 24779501]
32. Schaub MP, Wenger A, Berg O, Beck T, Stark L, Buehler E, et al. A Web-Based Self-Help Intervention With and Without Chat Counseling to Reduce Cannabis Use in Problematic Cannabis Users: Three-Arm Randomized Controlled Trial. J Med Internet Res 2015 Oct;17(10):e232 [FREE Full text] [doi: 10.2196/jmir.4860] [Medline: 26462848]
33. Myneni S, Lewis B, Singh T, Paiva K, Kim SM, Cebula AV, et al. Diabetes Self-Management in the Age of Social Media: Large-Scale Analysis of Peer Interactions Using Semiautomated Methods. JMIR Med Inform 2020 Jun 30;8(6):e18441 [FREE Full text] [doi: 10.2196/18441] [Medline: 32602843]
34. Barrera M, Glasgow R, McKay H, Boles S, Feil E. Do Internet-Based Support Interventions Change Perceptions of Social Support?: An Experimental Trial of Approaches for Supporting Diabetes Self-Management. Am J Community Psychol 2002;30(5):654. [doi: 10.1023/a:1016369114780]
35. Allen C, Vassilev I, Kennedy A, Rogers A. Long-Term Condition Self-Management Support in Online Communities: A Meta-Synthesis of Qualitative Papers. J Med Internet Res 2016 Mar 10;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]
36. PatientsLikeMe. URL: https://www.patientslikeme.com/ [accessed 2021-01-21]
37. BabyCenter. URL: https://www.babycenter.com/ [accessed 2021-01-21]
38. Fox S, Duggan M. Health Online 2013. The Pew Research Center. 2013. URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf [accessed 2021-01-05]
39. Mamykina L, Nakikj D. Collective Sensemaking in Online Health Forums. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems 2015 Apr:3217-3226. [doi: 10.1145/2702123.2702566]
40. Mankoff J, Kuksenok K, Kiesler S, Rode J, Waldman K. Competing online viewpoints and models of chronic illness. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Canada: ACM Press; 2011 May Presented at: CHI Conference on Human Factors in Computing Systems; May, 2011; Vancouver BC Canada p. 589-598. [doi: 10.1145/1978942.1979027]
41. Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. J Med Internet Res 2011;13(3):e62 [FREE Full text] [doi: 10.2196/jmir.1728] [Medline: 21846635]
42. Costa Figueiredo M, Caldeira C, Eikey EV, Mazmanian M, Chen Y. Engaging with Health Data: The Interplay Between Self-Tracking Activities and Emotions in Fertility Struggles. In: Proc. ACM Hum.-Comput. Interact. 2018 Nov Presented at: Proc ACM Hum-Comput Interact; 2018 Nov 1; - p. 1-20. [doi: 10.1145/3274309]
43. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res 2008;10(3):e15 [FREE Full text] [doi: 10.2196/jmir.1053] [Medline: 18504244]
44. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
45. Barker KK. Electronic support groups, patient-consumers, and medicalization: the case of contested illness. J Health Soc Behav 2008 Mar 01;49(1):20-36. [doi: 10.1177/002214650804900103] [Medline: 18418983]
46. Huh J, Ackerman MS. Collaborative Help in Chronic Disease Management: Supporting Individualized Problems. CSCW 2012;2012:853-862 [FREE Full text] [doi: 10.1145/2145204.2145331] [Medline: 25360442]
47. Gui X, Chen Y, Kou Y, Pine K, Chen Y. Investigating Support Seeking from Peers for Pregnancy in Online Health Communities. Proc. ACM Hum.-Comput. Interact 2017 Dec 06;1(CSCW):1-19. [doi: 10.1145/3134685]
48. Alqassim M, Kresnye K, Siek K, Wolters M. Facebook for Support versus Facebook for Research: The Case of Miscarriage. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. USA: ACM; 2019 May Presented at: CHI Conference on Human Factors in Computing Systems; May, 2019; Glasgow Scotland Uk p. 1-6. [doi: 10.1145/3290607.3312772]
49. Andalibi N. What Happens After Disclosing Stigmatized Experiences on Identified Social Media: Individual, Dyadic, and Social/Network Outcomes. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM Digital Library: ACM Press; 2019 May Presented at: CHI Conference on Human Factors in Computing Systems; May,

2019; Glasgow Scotland Uk p. 1-15 URL: https://dl.acm.org/doi/proceedings/10.1145/3290605 [doi: 10.1145/3290605.3300367]
50. D'Agostino AR, Optican AR, Sowles SJ, Krauss MJ, Escobar LK, Cavazos-Rehg PA. Social networking online to recover from opioid use disorder: A study of community interactions. Drug Alcohol Depend 2017 Dec 01;181:5-10. [doi: 10.1016/j.drugalcdep.2017.09.010] [Medline: 29024875]
51. Zhang M, Yang CC. Using content and network analysis to understand the social support exchange patterns and user behaviors of an online smoking cessation intervention program. J Assn Inf Sci Tec 2014 May 19;66(3):564-575. [doi: 10.1002/asi.23189]
52. Doherty K, Marcano-Belisario J, Cohn M, Mastellos N, Morrison C, Car J, et al. Engagement with Mental Health Screening on Mobile Devices: Results from an Antenatal Feasibility Study. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. New York: ACM Press; 2019 May Presented at: CHI Conference on Human Factors in Computing Systems; May, 2019; Glasgow Scotland Uk p. 1-15 URL: https://dl.acm.org/doi/10.1145/3290605.3300416 [doi: 10.1145/3290605.3300416]
53. MacLean D, Gupta S, Lembke A, Manning C, Heer J. Forum77: An Analysis of an Online Health Forum Dedicated to Addiction Recovery. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. Canada: ACM Press; 2015 Feb Presented at: Computer Supported Cooperative Work and Social Computing; March, 2015; Vancouver BC Canada p. 1511-1526. [doi: 10.1145/2675133.2675146]
54. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
55. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
56. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013 Sep;15(9):e189 [FREE Full text] [doi: 10.2196/jmir.2741] [Medline: 24014109]
57. Timeline of Selected FDA Activities and Significant Events Addressing Opioid Misuse and Abuse Internet. U.S. Food & Drug Administration. 2020. URL: https://www.fda.gov/media/126835/download [accessed 2021-01-05]
58. Gonzalez-Hernandez G. On the Ethics of Using Social Media Data for Health Research Internet. NLM Musings Mezzanine. 2019 Jun 25. URL: https://nlmdirector.nlm.nih.gov/2019/06/25/on-the-ethics-of-using-social-media-data-for-health-research/ [accessed 2019-09-11]
59. Barth S, de Jong MD. The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. Telematics and Informatics 2017 Nov;34(7):1038-1058. [doi: 10.1016/j.tele.2017.04.013]
60. Steinfeld N. "I agree to the terms and conditions": (How) do users read privacy policies online? An eye-tracking experiment. Computers in Human Behavior 2016 Feb;55:992-1000. [doi: 10.1016/j.chb.2015.09.038]
61. Liang OS, Sheffield JS, Taylor CO. Detecting Patterns of Prescription Drug Use During Pregnancy and Lactation with Visualization Techniques. In: AMIA Joint Summits on Translational Science proceedings. 2019 Presented at: AMIA Joint Summits on Translational Science; 2019 Mar 25; San Francisco, CA p. 478-487.
62. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
63. Wolff B, Mahoney F, Lohiniva A, Corkum M. Collecting and Analyzing Qualitative Data. In: Oxford Scholarship Online. CDC Field Epidemiol Man Internet Oxford, New York: Oxford University Press; 2018.
64. Hasin DS, O'Brien CP, Auriacombe M, Borges G, Bucholz K, Budney A, et al. DSM-5 criteria for substance use disorders: recommendations and rationale. Am J Psychiatry 2013 Aug;170(8):834-851 [FREE Full text] [doi: 10.1176/appi.ajp.2013.12060782] [Medline: 23903334]
65. HHS Guide for Clinicians on the Appropriate Dosage Reduction or Discontinuation of Long-Term Opioid Analgesics. U.S. Department of Health and Human Services. 2019. URL: https://www.hhs.gov/opioids/sites/default/files/2019-10/Dosage_Reduction_Discontinuation.pdf [accessed 2021-01-05]
66. Kummervold PE, Gammon D, Bergvik S, Johnsen JK, Hasvold T, Rosenvinge JH. Social support in a wired world: use of online mental health forums in Norway. Nord J Psychiatry 2002 Jul;56(1):59-65. [doi: 10.1080/08039480252803945] [Medline: 11869468]
67. Rubya S, Yarosh S. Video-Mediated Peer Support in an Online Community for Recovery from Substance Use Disorders. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Feb Presented at: Computer Supported Cooperative Work and Social Computing; 2017 Feb 25; Portland Oregon USA p. 1454-1469. [doi: 10.1145/2998181.2998246]
68. Bandura, Albert. Self-Efficacy: The Exercise of Control. In: Encyclopedia of Psychological Assessment. Thousand Oaks, California, United States: SAGE Publications Ltd; 1997:978.
69. Glanz K, Rimer B, Viswanath K. Health behavior and health education: Theory, Research, and Practice. In: Theory, Research, and Practice. 4th Edition. California: Jossey-Bass; 2008:978-0787996147.

70. Folkman S, Lazarus RS. The relationship between coping and emotion: implications for theory and research. Soc Sci Med 1988;26(3):309-317. [doi: 10.1016/0277-9536(88)90395-4] [Medline: 3279520]
71. Folkman S, Moskowitz JT. Positive affect and the other side of coping. American Psychologist 2000;55(6):647-654. [doi: 10.1037/0003-066x.55.6.647]
72. Practitioner and Program Data. Substance Abuse and Mental Health Services Administration. URL: https://www.samhsa.gov/medication-assisted-treatment/practitioner-program-data/certified-practitioners [accessed 2021-01-05]
73. Patrick SW, Buntin MB, Martin PR, Scott TA, Dupont W, Richards M, et al. Barriers to accessing treatment for pregnant women with opioid use disorder in Appalachian states. Subst Abus 2019 Oct 09;40(3):356-362. [doi: 10.1080/08897077.2018.1488336] [Medline: 29949454]
74. Nesvåg S, McKay JR. Feasibility and Effects of Digital Interventions to Support People in Recovery From Substance Use Disorders: Systematic Review. J Med Internet Res 2018 Aug 23;20(8):e255 [FREE Full text] [doi: 10.2196/jmir.9873] [Medline: 30139724]
75. Kim SJ, Marsch LA, Brunette MF, Dallery J. Harnessing Facebook for Smoking Reduction and Cessation Interventions: Facebook User Engagement and Social Support Predict Smoking Reduction. J Med Internet Res 2017 May 23;19(5):e168 [FREE Full text] [doi: 10.2196/jmir.6681] [Medline: 28536096]
76. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-Delivered Health Interventions That Work: Systematic Review of Meta-Analyses and Evaluation of Website Availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
77. Crane D, Garnett C, Brown J, West R, Michie S. Behavior change techniques in popular alcohol reduction apps: content analysis. J Med Internet Res 2015 May 14;17(5):e118 [FREE Full text] [doi: 10.2196/jmir.4060] [Medline: 25977135]
78. Wiebe R, Griffin A, Zheng Y, Harris K, Cleveland H. Twelve Steps, Two Factors: Coping Strategies Moderate the Association Between Craving and Daily 12-Step Use in a College Recovery Community. Subst Use Misuse 2018;53(1):127. [doi: 10.1080/10826084.2017.1325904] [Medline: 28813187]
79. Mason M. Sample Size and Saturation in PhD Studies Using Qualitative Interviews. Forum Qual Sozialforschung Forum Qual Soc Res Internet 2010:24.
80. Green BM, Van Horn KT, Gupte K, Evans M, Hayes S, Bhowmick A. Assessment of Adaptive Engagement and Support Model for People With Chronic Health Conditions in Online Health Communities: Combined Content Analysis. J Med Internet Res 2020 Jul 07;22(7):e17338 [FREE Full text] [doi: 10.2196/17338] [Medline: 32492651]

Abbreviations

CPS: Child Protective Service
DATA: Drug Addiction Treatment Act
MAT: medication-assisted treatment
NAS: neonatal abstinence syndrome
OHC: online health community
OUD: opioid use disorder
SUD: substance use disorder

Edited by G Eysenbach; submitted 17.02.20; peer-reviewed by S Li, JP Allem; comments to author 26.10.20; revised version received 11.12.20; accepted 20.12.20; published 04.02.21.

Please cite as:
Liang OS, Chen Y, Bennett DS, Yang CC
Identifying Self-Management Support Needs for Pregnant Women With Opioid Misuse in Online Health Communities: Mixed Methods Analysis of Web Posts
J Med Internet Res 2021;23(2):e18296
URL: https://www.jmir.org/2021/2/e18296
doi:10.2196/18296
PMID:33538695

Original Paper

Predictive Analytics for Care and Management of Patients With Acute Diseases: Deep Learning–Based Method to Predict Crucial Complication Phenotypes

Jessica Qiuhua Sheng[1]; Paul Jen-Hwa Hu[1], PhD; Xiao Liu[2], PhD; Ting-Shuo Huang[3], PhD, MD; Yu Hsien Chen[4], MD

[1]Department of Operations and Information Systems, David Eccles School of Business, University of Utah, Salt Lake City, UT, United States

[2]Department of Information Systems, WP Carey School of Business, Arizona State University, Phoenix, AZ, United States

[3]Department of General Surgery and Community Medicine Research Center, Keelung Chang Gung Memorial Hospital, Keelung, Taiwan

[4]Department of Chinese Medicine, College of Medicine, Chang Gung University, Taoyuan, Chang Gung, Taiwan

Corresponding Author:
Ting-Shuo Huang, PhD, MD
Department of General Surgery and Community Medicine Research Center
Keelung Chang Gung Memorial Hospital
No. 222, Maijin Road, Anle District
Keelung
Taiwan
Phone: 886 2 2431 3131
Email: huangts1234@gmail.com

Abstract

Background: Acute diseases present severe complications that develop rapidly, exhibit distinct phenotypes, and have profound effects on patient outcomes. Predictive analytics can enhance physicians' care and management of patients with acute diseases by predicting crucial complication phenotypes for a timely diagnosis and treatment. However, effective phenotype predictions require several challenges to be overcome. First, patient data collected in the early stages of an acute disease (eg, clinical data and laboratory results) are less informative for predicting phenotypic outcomes. Second, patient data are temporal and heterogeneous; for example, patients receive laboratory tests at different time intervals and frequencies. Third, imbalanced distributions of patient outcomes create additional complexity for predicting complication phenotypes.

Objective: To predict crucial complication phenotypes among patients with acute diseases, we propose a novel, deep learning–based method that uses recurrent neural network–based sequence embedding to represent disease progression while considering temporal heterogeneities in patient data. Our method incorporates a latent regulator to alleviate data insufficiency constraints by accounting for the underlying mechanisms that are not observed in patient data. The proposed method also includes cost-sensitive learning to address imbalanced outcome distributions in patient data for improved predictions.

Methods: From a major health care organization in Taiwan, we obtained a sample of 10,354 electronic health records that pertained to 6545 patients with peritonitis. The proposed method projects these temporal, heterogeneous, and clinical data into a substantially reduced feature space and then incorporates a latent regulator (latent parameter matrix) to obviate data insufficiencies and account for variations in phenotypic expressions. Moreover, our method employs cost-sensitive learning to further increase the predictive performance.

Results: We evaluated the efficacy of the proposed method for predicting two hepatic complication phenotypes in patients with peritonitis: acute hepatic encephalopathy and hepatorenal syndrome. The following three benchmark techniques were evaluated: temporal multiple measurement case-based reasoning (MMCBR), temporal short long-term memory (T-SLTM) networks, and time fusion convolutional neural network (CNN). For acute hepatic encephalopathy predictions, our method attained an area under the curve (AUC) value of 0.82, which outperforms temporal MMCBR by 64%, T-SLTM by 26%, and time fusion CNN by 26%. For hepatorenal syndrome predictions, our method achieved an AUC value of 0.64, which is 29% better than that of temporal MMCBR (0.54). Overall, the evaluation results show that the proposed method significantly outperforms all the benchmarks, as measured by recall, F-measure, and AUC while maintaining comparable precision values.

Conclusions: The proposed method learns a short-term temporal representation from patient data to predict complication phenotypes and offers greater predictive utilities than prevalent data-driven techniques. This method is generalizable and can be

applied to different acute disease (illness) scenarios that are characterized by insufficient patient clinical data availability, temporal heterogeneities, and imbalanced distributions of important patient outcomes.

(J Med Internet Res 2021;23(2):e18372) doi:10.2196/18372

KEYWORDS

data analytics; neural networks; phenotype; deep learning; electronic health records

Introduction

Background

Acute diseases and illnesses require timely and specialized care of patients whose conditions change rapidly, often within 48 hours of admission [1]. These diseases tend to evoke serious complications that develop quickly and can become fatal. Severe complications hinder patient recovery, substantially reduce their quality of life, create long-term impairments, and even cause death [2]. In general, a complication may have multiple subtypes or phenotypes, which signify and display distinct disease presentations [3,4]. Because phenotypes involve distinct symptoms and manifestations that require specific interventions, effective predictions of crucial complication phenotypes are crucial for physicians' timely diagnoses and therapeutic treatments to improve patient management and reduce mortality rates.

Several data-driven techniques aim at identifying phenotypic expressions from electronic health records (EHRs) and use them to predict important clinical events, such as complications [5]. Predictive analytics helps advance such data-driven approaches to predict complication phenotypes; however, this ability presents various challenges in acute disease scenarios for several reasons. First, in the early stages of an acute disease, essential clinical features and characteristics (eg, risk factors) associated with complication phenotypes may not be sufficiently available to predict the phenotypic outcomes. This data insufficiency constraint can greatly reduce the predictive utilities of data-driven techniques [6]. Second, patients undergo various laboratory tests, medical examinations, and therapeutic treatments, which are administered at different frequencies and time intervals. The resulting temporal heterogeneities (eg, pattern, time interval, frequency) create additional difficulties for phenotype predictions. Third, for any particular acute disease, crucial complication phenotypes may arise in a relatively small proportion of patients, further causing imbalanced distributions of patient outcomes.

Objectives

To address these challenges for effective complication phenotype predictions, we propose a novel recurrent neural network (RNN)-based method that incorporates a latent regulator (RNN-LR). Our method generates a temporal feature space representation with a recurrent neural network to cope with temporal heterogeneities in patients' conditions and disease progression and then uses the latent regulator to mitigate the data insufficiency constraint. We used a data set of 6545 patients with peritonitis to evaluate the ability of the proposed method to predict acute hepatic encephalopathy (AHE) [7] and hepatorenal syndrome (HRS) [8]—two crucial phenotypes of hepatic complications that can develop after surgical procedures for peritonitis (eg, laparotomy). Although only a small proportion of patients with peritonitis develop these phenotypes, they are life threatening and difficult to predict [7,9].

The following three benchmark methods are evaluated in our study: temporal multiple measurements case-based reasoning (MMCBR) [10], time-aware long short-term memory (T-LSTM) network [11], and time fusion convolutional neural network (CNN) [12]. The results show that the proposed method significantly outperforms all the benchmarks, as measured by recall, F-measure, and area under the curve (AUC) while maintaining comparable precision values. Although our illustrative evaluation focuses on complication phenotypes of peritonitis, the proposed method is generalizable and applicable to predict phenotypes of other acute diseases that are characterized by insufficient patient clinical data availability, temporal heterogeneities, and imbalanced distributions of patient outcomes.

Previous Work

Diseases can exhibit distinct phenotypic expressions [13]. For example, macrovascular disease spans six phenotypes, each associated with distinct anthropometric, clinical, and laboratory parameters [14]. Patients diagnosed with a particular disease can have complications pertinent to multiple phenotypes. Supported by accurate predictions of crucial complication phenotypes, physicians can improve their clinical decision making and patient management. To that end, predictive analytics empowers in-depth analyses of the rich patient clinical data in EHRs for the improved care and management of patients with acute diseases and illnesses [5,15].

Peritonitis

Peritonitis, an acute disease, is caused by the inflammation of the peritoneum [1] and often develops from bacterial or fungal infections [16]. Upon diagnosis with peritonitis, patients need immediate treatment (typically within 3 days), because it can progress rapidly and develop into life-threatening sepsis or septic shock [1]. Patients with peritonitis have higher mortality rates than those without peritonitis [17]. Several factors, such as age, sex, clinical conditions, and the living environment are associated with peritonitis-related mortality [17].

As two crucial phenotypic expressions of hepatic complications after peritonitis surgery, AHE and HRS can cause severe patient outcomes [7,18]. For patients with peritonitis who also have liver cirrhosis, intestinal bacterial overgrowth inside the body is responsible for hyperammonemia, which leads to AHE [19]. Similarly, HRS is a crucial complication phenotype of peritonitis with advanced cirrhosis too, that is characterized by renal failure and major disturbances in the circulatory function [8]. The

underlying mechanisms of HRS may result from complex changes in splanchnic and general circulation, as well as systemic and renal vasoconstrictors and vasodilators [20]. Both phenotypes are clinically important; however, AHE is more severe than HRS because it can deteriorate in a matter of hours.

In general, AHE is diagnosed by liver specialists, whereas HRS is diagnosed by liver specialists and nephrologists. Clinically, the determination of each phenotype depends on laboratory results and the patient's condition. Patients with AHE typically have hyperammonemia, hyperbilirubinemia, and central nervous system symptoms. Patients with HRS often display splanchnic arterial vasodilation and inflammation, which cause ascites and renal function impairment. Patients with AHE should receive lactulose and neomycin enema, which are infrequently used for other conditions according to the national health reimbursement policies in Taiwan. Thus, the occurrence of AHE can be assessed by the patient's condition. Patients with HRS are usually prescribed albumin and terlipressin, which can then be used to determine the occurrence of HRS. Each phenotype has a particular ICD-9 code: 572.2 for AHE and 572.4 for HRS.

The heterogeneity and variability in manifestations of hepatic encephalopathy among patients make it difficult to assess or predict patient conditions [21,22]. Previous studies have shown that AHE may be present in 50%-70% of patients with peritonitis who have cirrhosis, including those with abnormalities detectable only with psychometric testing [23]. The clinical manifestations of AHE include brain dysfunction and deep coma [7]. This phenotype represents a vital disease entity because the risk of dying within a year exceeds 60% after its development [24]. Furthermore, patients with spontaneous bacterial peritonitis have an estimated 30% chance of developing HRS [20]. Clinically, the only curative treatment for AHE and HRS is liver transplantation, but systemic infection is a contraindication to liver transplantation. Without timely detection and proper interventions, AHE and HRS can develop rapidly, create patient impairments, lead to life-threatening conditions, and have alarming mortality rates [7,9]. These phenotypes have unique clinical characteristics and features that can be analyzed with data-driven analytics for prediction. Overall, existing data-driven techniques for phenotype predictions can be classified as rule-based, machine learning–based, or deep learning–based. We review the representative studies of each in the upcoming sections.

Rule-Based Phenotype Predictions

Rule-based techniques [25-29] use clinically important features to depict the underlying phenotypes. A typical rule-based technique iteratively updates heuristic rules until its sensitivity and specificity satisfy the prespecified thresholds. Developing heuristic rules is labor intensive and time consuming because it requires iterative rule generation and substantial involvement from human experts. The prediction of disease phenotypes entails the extraction of clinically important features; essential features and their combinations in turn indicate the underlying disease phenotype. Guided by domain knowledge, previous research has developed heuristic rules to extract essential features (eg, medications, laboratory results, diagnoses) from EHRs for phenotype predictions, and then updated the extracted rules iteratively until sensitivity and specificity reached the prespecified levels. For example, the rule-based eMERGE technique uses EHRs, in combination with DNA biorepositories, to identify diabetic phenotypes and medication-induced liver lesions [29].

Machine Learning–Based Phenotype Predictions

Machine learning techniques coupled with EHRs can support and enhance the care and management of patients with peritonitis. For example, by integrating cellular and soluble biomarkers, support vector machines and tree-based algorithms can help physicians in predicting pathogen-specific immune responses of patients with peritonitis and guide them to formulate optimal antibiotic and operative therapies [30]. Previous research has also applied machine learning algorithms to predict phenotypes [5,28].

Existing machine learning–based techniques can be categorized as clustering analysis, graph-based learning, and probabilistic modeling. Techniques that rely on clustering analysis create phenotype clusters, such that patients in the same phenotypic cluster are more similar to one another than to patients in a different cluster. In essence, clustering analysis–based techniques [5,31] generate patient clusters so that patients with similar phenotypic expressions are in the same cluster. They usually use cross-sectional patient data to produce distinct clusters at a given time or analyze longitudinal clinical data to infer phenotypes that remain consistent over time [5]. However, existing clustering algorithms cannot deal robustly with high-dimensional patient data, and their applications are restricted to smaller, more homogenous data sets [5]. Most clustering-based techniques are applied to patient data at a single time point. Thus, clustering analyses of temporal data would require multiple applications of the chosen technique at different time points [31], further creating instability in the resulting phenotype clusters.

Graph-based techniques [32-34] can cope with temporal heterogeneities in longitudinal patient data (eg, pattern, time interval, frequency). They often assume sequential linkages of distinct clinical events and represent those events as temporally connected nodes in a graph [32]. However, this assumption does not always hold clinically. For example, patients frequently and concurrently receive multiple laboratory tests, treatments, or therapeutic procedures. Moreover, the graph construction process does not include laboratory results (values) that can be essential for inferring clinical outcomes [32]. In addition, probabilistic modeling can uncover the underlying phenotypes. For example, Pivovarov et al [35] propose UPhenome—an unsupervised, generative probabilistic model that can learn phenotypes from heterogeneous patient data. To identify chronic obstructive pulmonary disease subtypes that are similar in progression characteristics, Ross et al [36] develop a novel Bayesian nonparametric model that uses disease trajectory to represent the underlying biological or genetic similarity within the subtype.

The crucial peritonitis complication phenotypes that we study—AHE and HRS—can occur rapidly without any predictive signs, thereby hindering the use of conventional machine learning techniques for predictions. Recent

advancements in deep learning promise better predictions of patient outcomes [37,38] because they can learn from clinical sequences to account for complex patterns and relationships in sequential inputs. To illustrate, representation learning can extract complex relationships and nonlinearities among temporal events. Moreover, deep learning architectures, such as recurrent and convolutional neural networks, can be applied to better predict patient outcomes [39-41]. In the following sections, we review representative deep learning–based techniques that can deal with high-dimensional and temporally heterogeneous patient data.

Deep Learning–Based Phenotype Predictions

The use of predictive analytics for clinical decision support and patient management often involves large amounts of heterogeneous patient clinical data and needs to consider temporal relationships [42]. Fueled by fast-growing computational power and proliferating EHRs, deep learning has been applied in a broad array of diagnostic tasks, including those related to phenotypes [43,44]. For example, reconstructed RNNs with rectified linear units can impute missing values in genotype data to predict phenotype sequences [45]. Deep autoencoder techniques for unsupervised feature learning help clinicians in identifying acute leukemia phenotypes [46]. By combining latent representation learning of deep neural networks and causal inferences, Kale et al [40] discovered latent phenotypes that are causally predictive of clinical outcomes in patients in the intensive care unit. Moreover, deep RNNs can model multivariate clinical time series in a large data set and then transfer the knowledge to the limited labeled instances to classify the phenotypes of patients in the intensive care unit [47]. Existing literature suggests the value and feasibility of using deep learning in different diagnostic tasks and clinical contexts.

Particularly, EHRs contain rich, longitudinal patient clinical data that can be modeled as RNNs that can represent patients' records in an accurate and robust way [48]. These networks are effective for modeling patient (clinical) records as temporal logs of diagnostic results. For a particular patient, the state of disease or illness at time t is a summary of the diagnostic records before t. With each record represented as a feature vector, the vectors at different time points can provide sequential inputs to an RNN. The outputs at time t+1 can be used to produce a vector that represents the patient's state at t+1. Such patient-level vector representations can be further input into other (hidden) layers of the neural network to predict clinical outcomes (eg, readmission, mortality, complications). For predictive analytics in clinical scenarios, RNN-based deep learning architectures may be advantageous over traditional machine learning techniques. For example, an RNN can reduce or prevent adverse drug events by integrating heterogeneous, multidimensional drug data from different sources [49]. In addition, by coping with various clinical and temporal data, an uncertainty-aware convolutional RNN can predict patient mortality, with uncertainty denoting the irregular time intervals in patients [50].

Cost-Sensitive Learning

Many clinical diagnoses feature relatively few crucial cases among patients, which need to be properly addressed by data-driven techniques for prediction. If a sample has a substantially fewer number of minority class cases, standard classifiers generally cannot perform well because their predictions tend to steer toward the majority class. Cost-sensitive learning can address the imbalanced distributions of patient outcomes in a sample. It considers the misclassification cost (and possibly other costs) by assigning a high penalty (cost) to the misclassifications of a minority-class instance, without modifying the original data distribution in the sample [51]. Such learning essentially shifts the bias of a classification model in the favor of the minority class. By adjusting the costs associated with different misclassified labels [52], and with the goal of minimizing the total cost, cost-sensitive learning can produce greater predictive utilities. In many clinical scenarios, the minority class is relatively more important and has a higher misclassification cost. However, the overall performance of a classification model, whether machine learning– or deep learning–based, can be dominated by the majority-class instances. This issue may be addressed by combining evaluation results (eg, F-measure, AUC) and the costs associated with different outcome classes (eg, complication phenotypes) to optimize the cost parameter for effective classifications [51].

Research Gaps

This review of extant literature reveals several gaps. First, existing prediction techniques may be inadequate or ineffective for acute disease scenarios because previous phenotype research focuses largely on patients with chronic diseases [53,54], whose clinical conditions change less drastically than those of patients with acute diseases. In addition, patients with chronic diseases usually have fewer complications that develop rapidly and have more clinical data available for predictions compared with patients with acute diseases. Second, most previous research works [5,40,46,55] tend to overlook the data insufficiency constraint, which limits the use of early disease stage patient data to build effective computational models for predicting complication phenotypes. Several studies have identified disease phenotypes by assuming full patient data availability [40,46]; however, clinical data captured in the early stages of an acute disease may lack essential information for predicting complication phenotypes. Some important clinical characteristics and factors of complication phenotypes may be available in the early stages but are not sufficiently informative for predicting phenotypic outcomes. Third, complication phenotypes associated with an acute disease often have an imbalanced distribution of different outcomes. Fourth, the clinical efficacy of data-driven techniques for complication phenotype predictions still requires adequate empirical evaluations, especially in acute disease situations that feature data insufficiencies and imbalanced distributions of patient outcomes.

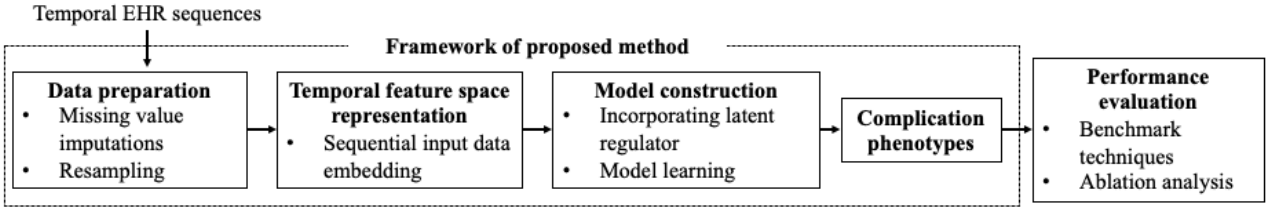
Effective complication phenotype predictions need to address these challenges and consider patients' heterogeneities and disease progression variations over time while coping with the data sufficiency constraint. We propose a deep learning–based method that leverages temporal feature space representation to address temporal heterogeneities in patient data. Although previous research works have acknowledged the importance of unobserved latent factors for influencing phenotypes [40,56], few studies have explicitly considered such factors for

phenotype predictions. As a remedy, we incorporate a latent parameter matrix to account for unobserved (subsequent) patient condition and disease progression variations. In addition, our method addresses missing values in patient data and includes cost-sensitive learning, which can address imbalanced outcome distributions by combining evaluation results (F-measure and AUC) and the cost associated with each complication phenotype to optimize cost parameters for an improved predictive performance.

Methods

We elaborate on the proposed method in Figure 1. As shown in the figure, this method involves data preparation, temporal feature space representation, model construction, and model evaluation.

Figure 1. Overall processing of recurrent neural network-latent regulator. EHR: electronic health record.



Data Preparation

Missing data prevail in many clinical scenarios and create fundamental challenges for predictive analytics [57]. Patients with acute diseases are often closely monitored with various laboratory tests, but missing values arise when the test results are not properly and consistently recorded because of the physician's preference, recording errors, or other reasons. For data preparation, we perform expert-guided feature selection to identify the clinical attributes and laboratory tests that are essential to a severe complication and then employ a deep autoencoder-based model to impute missing values for these features. The deep autoencoder model [58] identifies patients similar to the focal patient and uses their attribute values to infer and replace the patient's missing values [59]. Because only a relatively small proportion of patients may develop severe complications, we apply the SMOTEENN (Synthetic Minority Oversampling Technique–Edited Nearest Neighbors) algorithm [60] to address the imbalanced distributions of different outcome classes.

Temporal Feature Space Representation

Patient data, including vital signs and laboratory results, are longitudinal and pertain to different clinical events over time. A clinical sequence reflects the patient's disease progression and has heterogeneous characteristics that may prevent clinically actionable insights. To extract acute disease progression from sequential (clinical) events, we apply sequence embedding, which is a feature leaning technique that projects sequential events into vectors of numeric numbers. In general, patients sharing similar clinical conditions are closer in distance than otherwise. Therefore, we used a temporal representation to depict each patient's disease progression. We assume that a patient p has a set of temporal clinical events that occur between time and . At each time point, a patient may have multiple temporal features (eg, test results, diagnoses), denoted by , where m indicates the number of diagnosis categories at . In addition, each patient has demographic data (eg, sex, age), represented as , where P is the total number of patients. With observed clinical (event) sequences, we can construct a temporal feature representation () for that patient:

Where is a function that can project the temporal clinical sequences to the temporal feature representation ().

Model Construction

Variations that exist in patients' conditions and disease progression cannot be fully explained by patients' demographics, laboratory results, and therapeutic (surgical) data [61]. Therefore, we include an additional parameter matrix , which serves as a latent regulator to account for disease progression information or underlying mechanisms related to complication phenotypes. In addition, refers to the disease progression space and comprises information extracted from clinical data available in early disease stages; that is, is the temporal feature space extracted from patients' clinical data. The data available in the early stages of an acute disease are usually limited and cannot reveal a patient's subsequent progression or effectively predict complication phenotypes. To alleviate this constraint, acts as a latent regulator, independent of the disease progression space () to account for unobserved variations in the subsequent patient condition and disease progression.

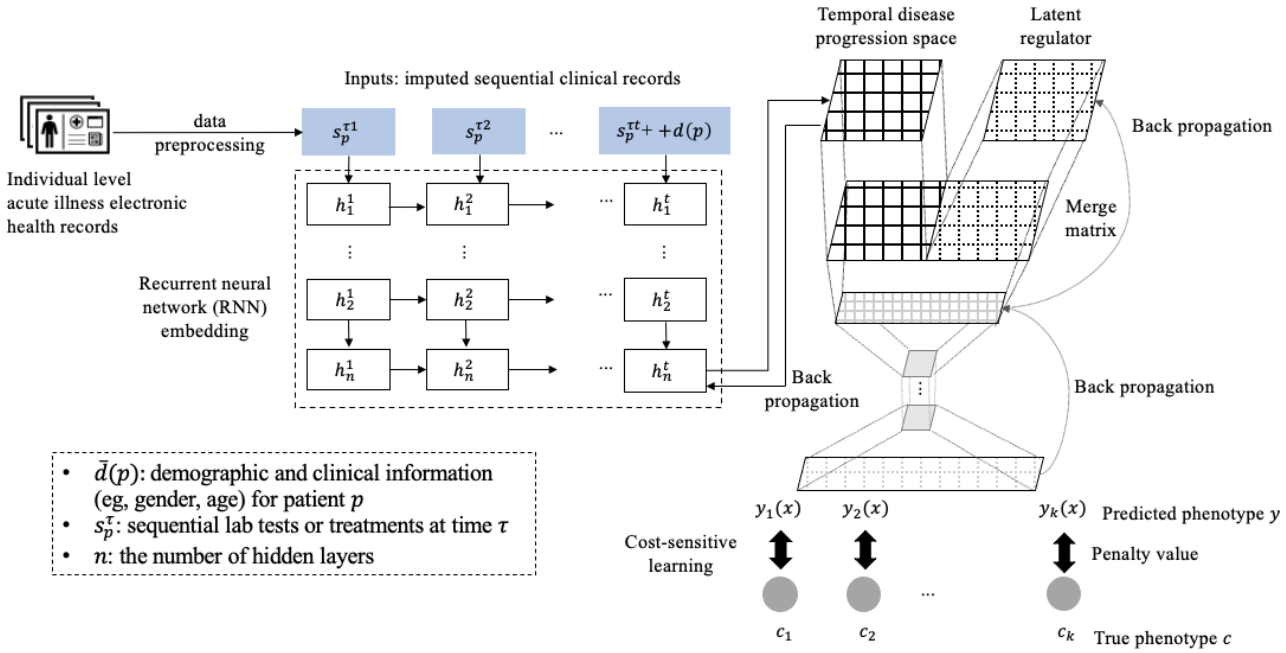
We assume a combined effect of and , for which is generalized with iterative clinical feature updates, according to

where γ is the learning rate, and indicates the minimum number of iterations required to converge . During model

learning, gradually converges to a stable range, as depicted in . For testing, the parameter matrix () facilitates phenotypic predictions for individual patients.

Although severe complications represent the minority class in patient data, they have profound effects on patient outcomes and health care costs. Hence, we employ cost-sensitive learning to better predict the minority class, according to the respective misclassification costs, by applying the cost matrix to penalize incorrect predictions (misclassifications). Figure 2 presents the proposed RNN-LR method.

Figure 2. Proposed latent regulator-embedded recurrent neural network method.



Our method minimizes the expected costs of incorrect phenotype predictions, calculated as:

where is the learning function of and reveals the combined effect of disease progression and a latent regulator on prediction, K denotes the total number of classes, $c(k, i)$ indicates the cost of misclassifying an instance of class k as class i. estimates the probability of class i, given (). As shown in Figure 2, and serve as input to the second multilayer neural network and thereby are mapped into the phenotype space. We use a SoftMax function to estimate the probability that an instance is classified as each distinct outcome class. In the output layer, the ith node contains weight () and bias (). For each phenotype outcome class, the probability of phenotype i, given can be calculated as:

Finally, we employ cross-entropy as a loss function to learn and optimize the model parameters:

Thus, parameters get updated through Adam optimizer and back propagation.

Data

Data Sources

We obtained a clinical data set from the Department of Laboratory Medicine at the Chang Gung Memorial Hospital, which is accredited by the Taiwan Accreditation Foundation and approved by the American College of Pathologists [62,63]. This data set consists of 10,354 records pertaining to 6545 patients who underwent peritonitis surgery between 2003 and 2015. Designated professionals at the hospital integrated patient records in EHRs, according to the common format of the Chang Gung Research Database (CGRD) that provides standardization and facilitates data extraction, transformation, and loading for analyses [64].

The CGRD links directly to the National Health Insurance Research Database (NHIRD), which informs reimbursement decisions [62,63]. To prevent fraud and contain costs, the National Health Insurance Administration of the Ministry of Health and Welfare performs frequent, random audits. Thus, the data in the NHIRD and CGRD are reliable and accurate. Information systems professionals at the hospital also assessed the data conversion process and integrity to ensure that all patient records (including diagnoses and laboratory results) are correctly transferred to the CGRD without errors or losses [64]. Experienced personnel from the Research Institute of Chang Gung Memorial Hospital assisted us in compiling the data set. In particular, they performed data preprocessing and consolidation to ensure that each patient's records were collected within the same time interval.

Data Processing and Details

In the data set, each laboratory test has a time stamp that indicates when the results are available (reported). We used the date of peritonitis surgery as the starting point and collected patient data over the next 3 days. This procedure ensured that all patient records were collected within the same time interval after surgery. Each patient had at least one clinical record within the three-day window. Specifically, 3665 patients had 1 clinical record, 1951 patients had 2, and 929 patients had 3 clinical records. Each record contains the patient's demographic and clinical data (eg, age, sex, comorbidity) and potentially multiple laboratory results, which we used to predict the complication phenotypes (AHE, HRS, or neither).

Table 1 summarizes the patient demographic and clinical variables in the data set. We adopted the International Classification of Diseases, Ninth Revision, Clinical Modification comorbidity coding algorithm [65] to define the Charlson comorbidities in administrative and clinical data. Accordingly, the comorbidity types in our study span 17 categories, from mild liver disease to moderate or severe liver disease to myocardial infarction. Among the 6545 peritonitis patients in the data set, 41 developed AHE, and 174 developed HRS, that is, the distribution of different outcome classes (AHE, HRS, or neither) is highly imbalanced. Furthermore, AHE and HRS have profound implications for patient outcomes and mortality. For example, 9 of the 41 patients with AHE in the sample died, with a mortality rate of 22%; patients with AHE or HRS had an average length of stay of 25 days (hospitalization), whereas patients without these phenotypes had an average of 17 days.

Table 1. Summary of patient demographic and clinical variables in the peritonitis data set.

Variables	Type	Value or range	Description
Variable name			
Patient_ID	String	Unique patient ID.	Patient ID
Inpatient_ID	String	Unique inpatient ID	Inpatient ID
Sex	Integer	0 or 1	Male or female
Age (<20, 20-60, or >60 years)	Integer	(0-102)	Patient's age
Inpatient variables			
DGTM[a]	Integer	(0-136)	The number of hospitalizations before peritonitis surgery
Comorbidities (including malignant tumor)	String	[Mild liver disease, …, Renal disease]	17 classified comorbidities that include liver disease, according to ICD-9[b] coding algorithms [65]
Operation category	String	0, 1, 2, 3, 4, 5, or 6	Seven categories in total (adopted from ICD-9-CM[c]): cholecystitis or cholangitis, appendicitis, hollow organ perforation, bowel ischemia, intestinal obstruction, hernia with bowel gangrene, hernia with bowel obstruction
Complication pneumonia	Integer	0 or 1	Whether a patient has pneumonia
Complication UTI[d]	Integer	0 or 1	Whether a patient has UTI
Complication SSI[e]	Integer	0 or 1	Whether a patient has SSI
Laboratory test results			
Albumin	Float	(0.19-5.5g per dL)	Serum albumin level
Amylase	Float	(5-11873 U per L)	Serum amylase
BNP[f]	Float	(21.49-4942 pg per ml)	Levels of B-type natriuretic peptide
BUN[g]	Float	(1-281 mmol per L)	Blood urea nitrogen
Band[h]	Float	(0.3-76)	Band neutrophil
Ca[i]	Float	(0.8-14.49 mg per dL)	Calcium in blood
CR[j]	Float	(0.07-34.2 mg per dL)	Serum creatinine
CRP[k]	Float	(0.2-679.2 mg per L)	C-reactive protein test
Hematocrit	Float	(1-63.2%)	Hematocrit test (proportion of red blood cells in the blood)
INR[l]	Float	(0.79-12)	International normalized ratio of prothrombin time
K[m]	Float	(1.60-18.86 mEq per L)	Blood potassium test
Lactate	Float	(5.49-240.6 mmol per L)	Level of lactic acid
Leukocyte (WBC[n])	Float	(0.2-141.5×10^9 per L)	White blood cell count
Platelets	Float	(0.7-1373×10^9 per L)	The number of platelets in blood
Prealbumin	Float	(2-44.59 mg per dL)	Prealbumin in blood
Procalcitonin	Float	(4.56-45.49 μg per L)	Blood test to diagnose sepsis
Total bilirubin	Float	(0.1-56.54 mg per dL)	Total amount of bilirubin in blood
Na[o]	Float	(10-190 mEq per L)	Blood sodium test
Target variable (crucial complication phenotype to be predicted)			
Complication phenotype	Categorical	AHE[p], HRS[q], or Neither	

[a]DGTM: number of hospitalization before peritonitis surgery.

[b]ICD-9: International Classification of Diseases, Ninth Revision.

[c]ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification.
[d]UTI: urinary tract infection.
[e]SSI: surgical site infection.
[f]BNP: B-type natriuretic peptide.
[g]BUN: blood urea nitrogen.
[h]Band: bandemia.
[i]Ca: calcium.
[j]CR: creatinine.
[k]CRP: C-reactive protein.
[l]INR: international normalized ratio.
[m]K: potassium.
[n]WBC: white blood cell.
[o]Na: sodium.
[p]AHE: acute hepatic encephalopathy.
[q]HRS: hepatorenal syndrome.

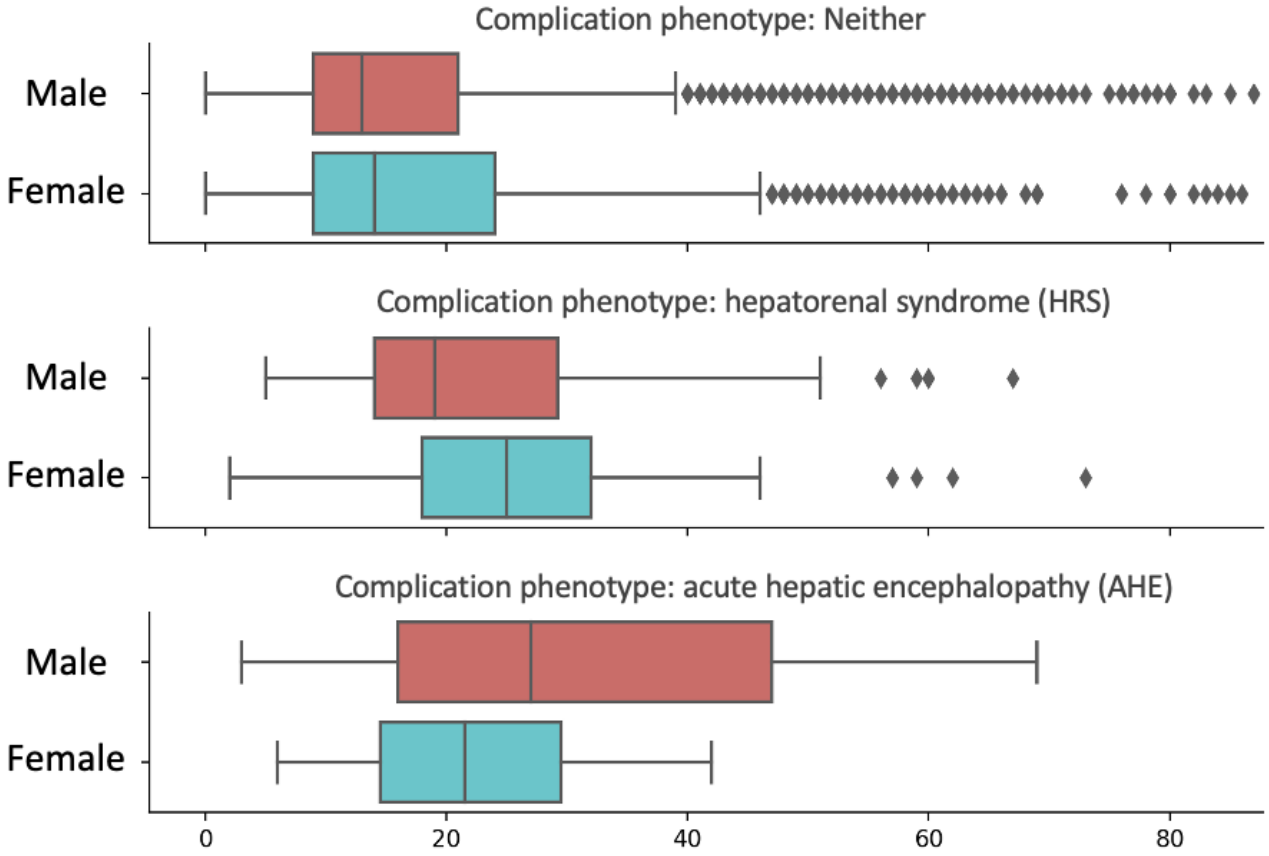
For the 41 patients with AHE in the sample, the complication phenotype occurred between 6 and 17 days after peritonitis surgery, with a mean of 13 days. For the 147 patients with HRS, the complication phenotype developed between 4 and 47 days after the surgery, with a mean of 14 days. In their clinical trial, Huang et al [66] reported a survival curve that indicated that most patients discontinued antibiotic treatment within 3 days after peritonitis surgery, suggesting that AHE and HRS seldom occur within those 3 days. In general, complication phenotypes, including AHE and HRS, arise approximately 2 or 3 weeks after peritonitis surgery [66,67]. For example, HRS is characterized by a rapid, progressive impairment of renal function. Furthermore, it develops, on average, about 2-3 weeks after peritonitis surgery [68,69]. Similarly, patients develop AHE approximately 2 weeks after peritonitis surgery [70,71]. The combined evidence from the relevant literature and clinical findings indicates the appropriateness of using the data available 3 days after surgery to predict subsequent AHE and HRS occurrences.

Descriptive Statistics

We provide some descriptive statistics related to gender, the number of hospitalizations before peritonitis surgery, and different complication phenotypes (ie, AHE, HRS, and neither) in the data set. The average number of hospitalizations before surgery was slightly lower for AHE and HRS than for neither: 2.6 for AHE, 2.4 for HRS, and 2.7 for neither. Both male and female patients had a similar number of hospitalizations: approximately 2.5 times. Female patients with neither of the diseases had more hospitalizations (about 3.2 times) than their male counterparts. These differences in part reflect the risk: both AHE and HRS may arise abruptly, even without many previous hospitalizations.

We also analyzed the relationships of sex, complication phenotypes, and the length of stay after surgery. As shown in Figure 3, both AHE and HRS induce longer lengths (20 days+) after surgery, whereas for neither, the length of stay was approximately 17 days. The longer length of stay associated with AHE and HRS again underscores the importance of phenotype predictions. For AHE, male patients had a longer length of stay than female patients, but we observed an opposite pattern for HRS.

Figure 3. Analysis of sex, complication phenotype, and the length of stay in the data set.



Evaluation Design

Benchmark Techniques

In total, three prevalent techniques were included in the evaluation as benchmarks: temporal MMCBR [10], time fusion CNN [12], and T-LSTM [11]. First, temporal MMCBR performs clustering analyses to identify similar (patient) temporal sequences in a sample [10]; therefore, it can handle temporal patient data that vary in their time intervals and granularity. Second, time fusion CNN, a deep learning–based technique, learns patient representations and measures pairwise similarity in temporal patient data to capture important characteristics specific to individual patients [12]. Third, patient subtyping through T-LSTM, another deep learning–based technique, can cope with patient data that feature temporal heterogeneities by employing autoencoders to learn patient representations, which helps cluster patients into subtypes [11]. Our benchmarks do not include graph-based techniques because concurrent laboratory tests in the data set make them inadequate for representing patient conditions and disease progression in a 2D graph. We also exclude probabilistic modeling that offers limited predictive utilities in situations involving imbalanced samples. In the evaluation, the proposed method and all benchmark techniques employed the same cost-sensitive matrix.

Implementation and Parameter Tuning

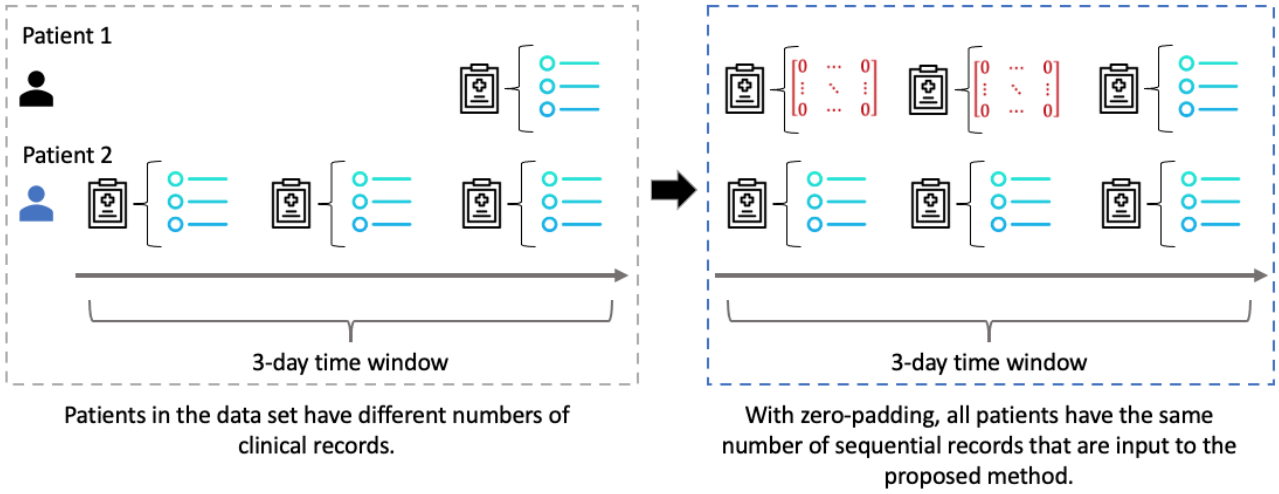
All the evaluations were performed on a computer with a dual-core processor of 2.7 GHz and 8 GB of memory, running macOS Catalina. We used the SMOTEENN algorithm from the Python imbalanced-learn library and applied the Python Shapley Addictive Explanations (SHAP) package to obtain SHAP values for the feature importance analysis. The proposed method and benchmark techniques were implemented using PyTorch. Our method constructs an 8-layer RNN to map the disease progression space with a latent regulator and adopts a multilayer perceptron neural network with three dense layers to predict complication phenotypes. Specifically, the RNN embedding produces a 2D vector, 8×8 in size, which depicts the temporal disease progression space. We randomly split the data set into 80% for training and 20% for testing. The testing set had 12 AHE cases, 35 HRS cases, and 1262 neither cases. For misclassified labels, we set the initial cost parameter for each phenotype (AHE or HRS) to 200, in line with a related research [72]. We performed a series of parameter tuning analyses, and then used the results to determine essential hyperparameter values (Table 2), including an optimal number of layers for each neural network.

Table 2. Essential hyperparameters used in the proposed method.

Parameter	Value
Learning rate (γ)	0.01
Latent regulator size	8×8
Drop rate	0.2
Number of layers	8
Units in each layer	16
Weight decay (λ)	0.005

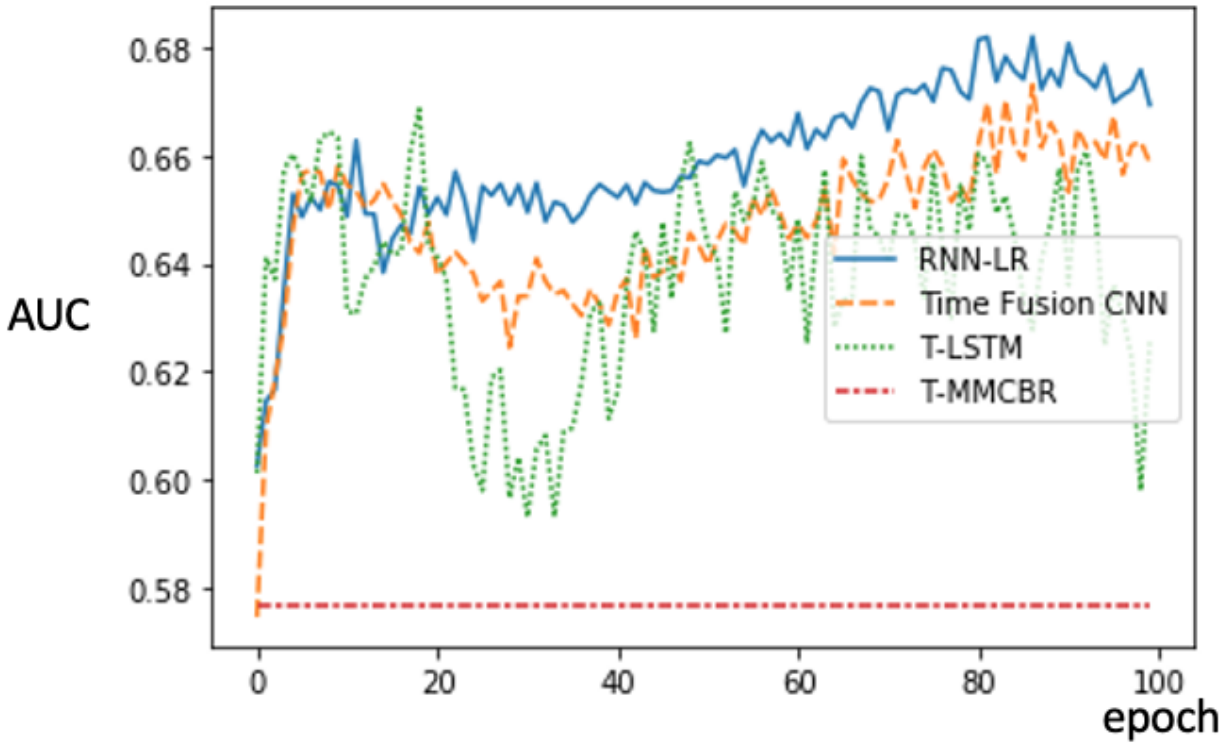
A clinical record contains the results of the laboratory tests prescribed by the physician, that is, a record has one timestamp. As noted, patients in the data set have different numbers of records within the 3-day window after peritonitis surgery, which are used for model construction. Because the proposed RNN-LR method requires the same number of clinical records for each patient, we employed zero padding to ensure that each patient had three sequential records. As a result, the data set contains a total of 19,635 clinical records: 6545 patients × 3 (sequential) records. We illustrate the zero-padding process in Figure 4. If a patient has only one clinical record within the 3-day window, we place that record at the end of the sequence and fill the first 2 records with zeroes, according to the length of the longest sequence (three). Hence, our method uses input sequences of the same length for model training.

Figure 4. Use of zero padding to prepare clinical records for the proposed recurrent neural network-latent regulator method.



We used the ReLU for activation and the cross-entropy function for optimization. Finally, the Adam optimizer was applied to update the model parameters. Figure 5 presents the learning curves of the proposed method versus benchmark techniques [73]. As shown, the hyperparameters appear to converge toward optimality after 100 epochs. Notably, our method consistently achieves a greater AUC than the benchmark techniques after 60 epochs.

Figure 5. Analysis of learning curves of the proposed method versus benchmark techniques. AUC: area under the curve; CNN: convolutional neural network; MMCBR: multiple measurement case-based reasoning; RNN-LR: recurrent neural network-latent regulator; T-LSTM: time aware long short-term memory.



Performance Measures

We evaluated predictive performance in terms of recall, precision, F-measure, and AUC. In line with previous research works [74,75], we adopted the one-against-all strategy to examine the respective techniques across different performance measures and outcome classes. To illustrate, we combine HRS and neither as a single class to calculate the precision, recall, and F-measure while evaluating AHE predictions. This approach reduces 3 outcome classes to 2 (AHE and no AHE), with AHE as the positive class and no AHE as the negative class. Similarly, we consider HRS as the positive class and no HRS as the negative class while assessing HRS predictions. Recall, or sensitivity, indicates the fraction of correctly predicted positive observations among all actual positive-class observations, calculated as:

Precision, or positive predictive value, denotes the ratio between the correctly predicted positive (or negative) observations and the total predicted positive (or negative) observations, calculated as:

A high recall value reflects the ability to predict patients who will develop AHE or HRS, whereas a high precision value signals a low false positive rate. The F-measure is the harmonic mean of precision and recall, with 1 indicating the best performance and 0 indicating the worst [76], calculated as:

Finally, the AUC depicts a technique's overall ability to distinguish different outcome classes across various threshold values. Because we apply the one-against-all strategy in the evaluation, the AUC reveals a technique's performance relative to a random classification, without any biases associated with the sample size used in the evaluation.

Results

Imputation Performance

We evaluated the performance of several prevalent missing value imputation techniques: multivariate imputation by chained equations [77], SoftImput [78], a K-nearest neighbors technique [79], and a deep autoencoder model [58]. To compare their effectiveness, we randomly removed 20% and 30% of the laboratory results from the data set, applied each technique to impute the missing values, and then calculated the normalized root mean squared error between the predicted and holdout values. The NRMSE is the difference between the imputed and the holdout values divided by the average value of the complete data. As shown in Table 3, the deep autoencoder model, which we incorporated into the proposed method, consistently exhibits the best imputation performance consistently.

Table 3. Missing value imputation performance of respective techniques.

Technique (reference)	NRMSE[a]	
	20% imputation performance	30% imputation performance
MICE[b] [77]	0.8479	0.8886
SoftImpute [78]	0.8546	0.9044
KNN[c]-based imputation [79]	1.0209	1.0044
Deep autoencoder model [58]	0.7926	0.8453

[a]NRMSE: normalized root mean squared error.

[b]MICE: multivariate imputation by chained equations.

[c]KNN: K-nearest neighbors.

Prediction Performance Evaluation

Table 4 presents the results of prediction evaluation. Overall, the proposed method outperforms all the benchmarks for predicting AHE and HRS, as measured by recall, F-measure, and AUC. Because recall indicates the ability to identify patients who are likely to develop AHE or HRS, it is arguably more important than precision. For predicting AHE, our method achieves 27%-147% improvements in recall, 26%-64% in AUC, and 56%-100% in the F-measure compared with the benchmarks. For HRS predictions, we observed 5%-300% improvements in recall, up to 19% improvement in AUC, and up to 30% in the F-measure. The recall level achieved by our method, recorded at 0.42 for AHE and 0.40 for HRS, is significantly higher than that of the best-performing benchmark (T-LSTM). Similarly, the AUC values attained by the proposed method, 0.82 for AHE and 0.64 for HRS, are significantly greater than those of T-LSTM or time fusion CNN. Jointly, the F-measure and AUC values attained by the proposed method indicate its greater effectiveness in predicting the crucial complication phenotypes than the benchmark techniques because of its high recall and comparable precision. Together, these results reveal that the proposed method can help physicians concentrate on patients who are more likely to develop severe complications.

Table 4. Prediction performance of the proposed method versus benchmark techniques.

Technique and outcome class	Precision	Recall	F-measure	AUC[a]
Temporal MMCBR[b]				
AHE[c]	0.20	0.17	0.18	0.50
HRS[d]	0.09	0.10	0.10	0.54
Neither	0.97	0.96	0.96	0.54
Time fusion CNN[e]				
AHE	0.12	0.33	0.18	0.65
HRS	0.06	0.38	0.10	0.67
Neither	0.97	0.81	0.88	0.63
T-LSTM[f]				
AHE	0.09	0.33	0.14	0.65
HRS	0.05	0.27	0.09	0.68
Neither	0.97	0.84	0.90	0.63
RNN-LR[g]				
AHE	0.21	0.42	0.28	0.82
HRS	0.08	0.40	0.13	0.64
Neither	0.98	0.85	0.91	0.66

[a]AUC: area under the curve.

[b]MMCBR: multiple measurement case-based reasoning.

[c]AHE: acute hepatic encephalopathy.

[d]HRS: hepatorenal syndrome.

[e]CNN: convolutional neural network.

[f]T-LSTM: time aware long short-term memory.

[g]RNN-LR: recurrent neural network-latent regulator.

We performed paired *t* tests to examine whether the performance improvements achieved by our method over benchmark techniques are significant. Specifically, we independently evaluated each technique 100 times, and then used the results to test the significance of the performance differentials. As shown in Figure 6, the improvements in the weighted F-measure and AUC associated with our proposed method are statistically significant at $P<.05$. Table 5 details the paired *t* test results.

Figure 6. Predictive performances of our method versus benchmarks, 100 evaluation trials. AUC: area under the curve; CNN: convolutional neural network; MMCBR: multiple measurement case-based reasoning; RNN-LR: recurrent neural network-latent regulator; T-LSTM: time aware long short-term memory.

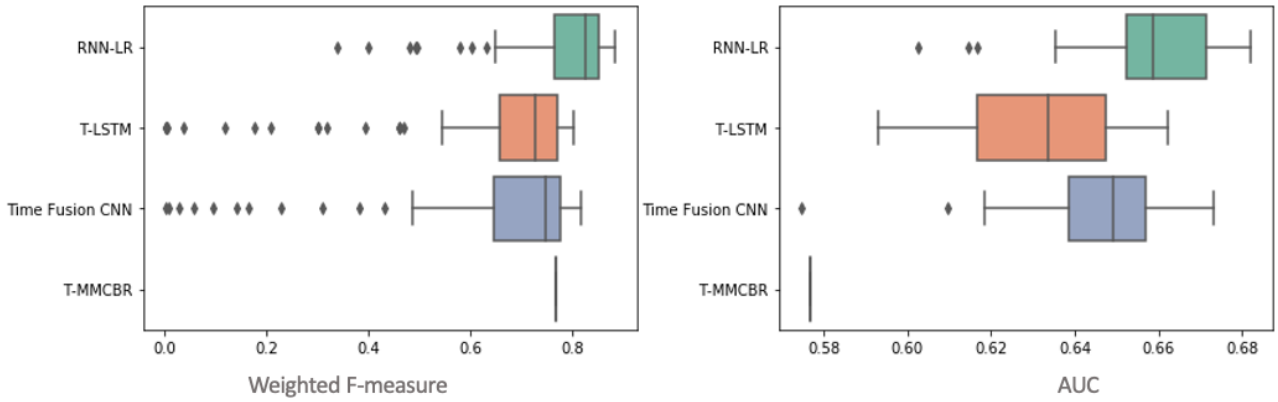


Table 5. Paired *t* test results of improvements by our method over benchmark techniques.

Proposed method (RNN-LR[a])	AUC[b]			Weighted F-measure		
	T-LSTM[c]	Time fusion CNN[d]	Temporal MMCBR[e]	T-LSTM	Time fusion CNN	Temporal MMCBR
P value	<.001	<.001	<.001	<.001	<.001	<.05

[a]RNN-LR: recurrent neural network-latent regulator.

[b]AUC: area under the curve.

[c]T-LSTM: time aware long short-term memory.

[d]CNN: convolutional neural network.

[e]MMCBR: multiple measurement case-based reasoning.

Table 6 presents the confusion matrix created by testing the case predictions with our method. Because neither account for the majority of the peritonitis sample, we observe a tendency to predict AHE or HRS as neither.

Table 6. Confusion matrix of test sample predictions by the proposed method.

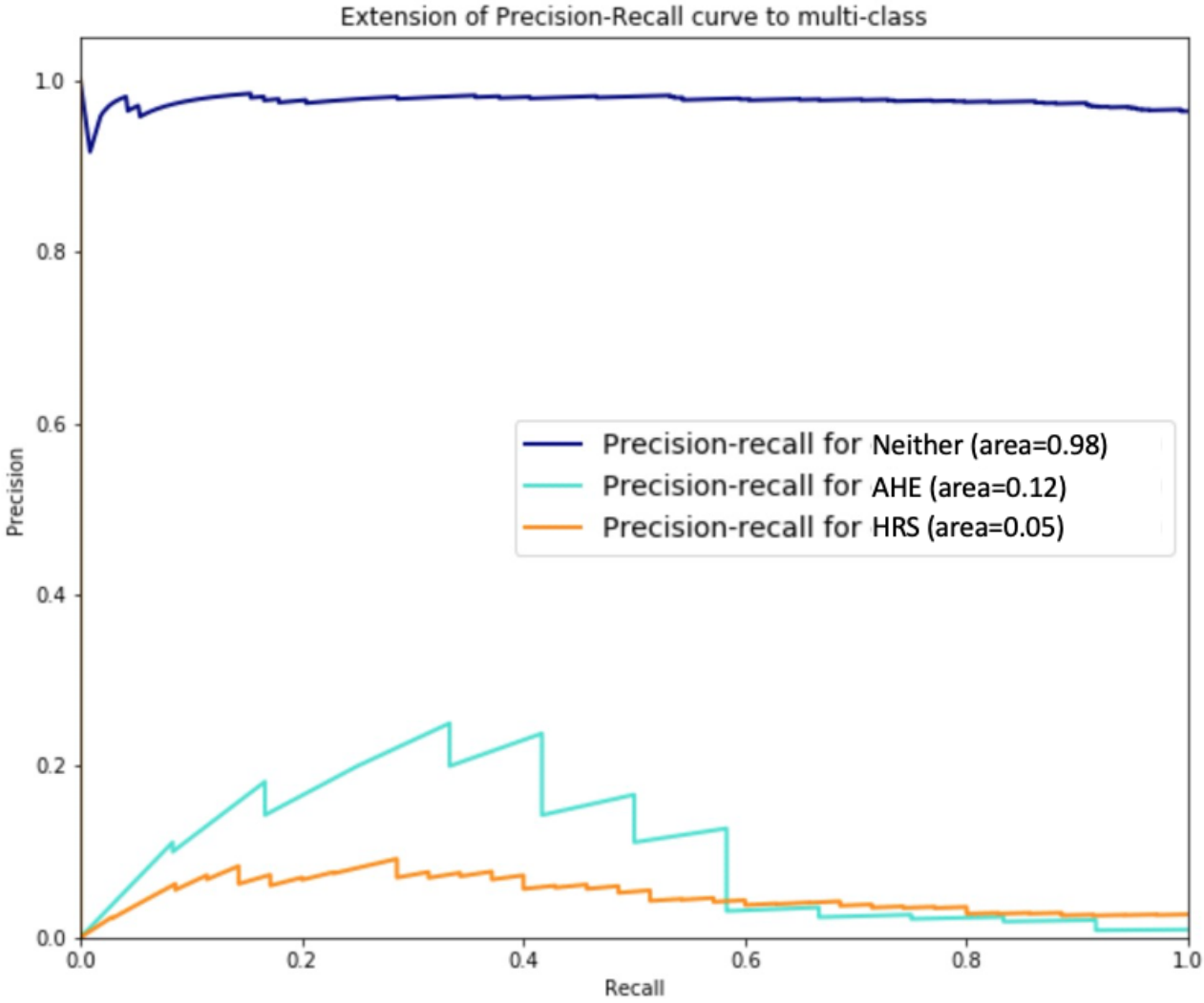
Actual outcome class	Predicted outcome class		
	Predicted neither	Predicted AHE[a]	Predicted HRS[b]
Actual neither	1079	18	165
Actual AHE	7	5	0
Actual HRS	20	1	14

[a]AHE: acute hepatic encephalopathy.

[b]HRS: hepatorenal syndrome.

The relatively low precision values of both our proposed method and the benchmark techniques can be attributed to the highly imbalanced outcome class distributions: AHE and HRS cases only account for 0.6% and 1.1% of the sample, respectively. Figure 7 depicts the precision-recall curves that reveal their trade-off across different thresholds. Although both AHE and HRS have low precision and recall values because of their imbalanced distributions in the sample, a higher recall value for each phenotype could be achieved at the cost of a lower precision rate. Because AHE cases account for 0.6% of the sample, our method, in the best scenario, can correctly predict 20% of AHE cases, hence representing a substantial improvement over random guessing. For both AHE and HRS, the low AUC values do not necessarily convey poor performance; rather, they indicate that the imbalanced distributions make accurate phenotype predictions very difficult.

Figure 7. Precision-recall curves for the proposed method regarding acute hepatic encephalopathy, hepatorenal syndrome, and neither of these conditions.



Value of Latent Regulator to Predictive Power

We incorporate a latent regulator in the RNNs as an important novelty of the proposed method. We therefore specifically examined its value regarding our proposed method's predictive utilities. Table 7 summarizes the model's predictive power, with and without the latent regulator. The method is more effective for predicting AHE if it includes the latent regulator, as indicated by the 40% improvement in recall, 24% in AUC, 31% in precision, and 33% in the F-measure, compared with its application without the regulator. For HRS, incorporating the latent regulator leads to an improvement of 7% in AUC, 14% in precision, and 18% in F-measure. These comparative results confirm that the value of the latent regulator (latent parameter matrix) is relative to techniques that only rely on available patient data.

Table 7. Predictive performance of the proposed method, with versus without the latent regulator.

Performance measure	Our method without latent regulator			Our method (RNN-LR[a])		
	Neither	AHE[b]	HRS[c]	Neither	AHE	HRS
Precision	0.97	0.16	0.07	0.98	0.21	0.08
Recall	0.89	0.30	0.26	0.85	0.42	0.40
F-measure	0.93	0.21	0.11	0.91	0.28	0.13
AUC[d]	0.58	0.66	0.60	0.66	0.82	0.64

[a]RNN-LR: recurrent neural network-latent regulator.

[b]AHE: acute hepatic encephalopathy.

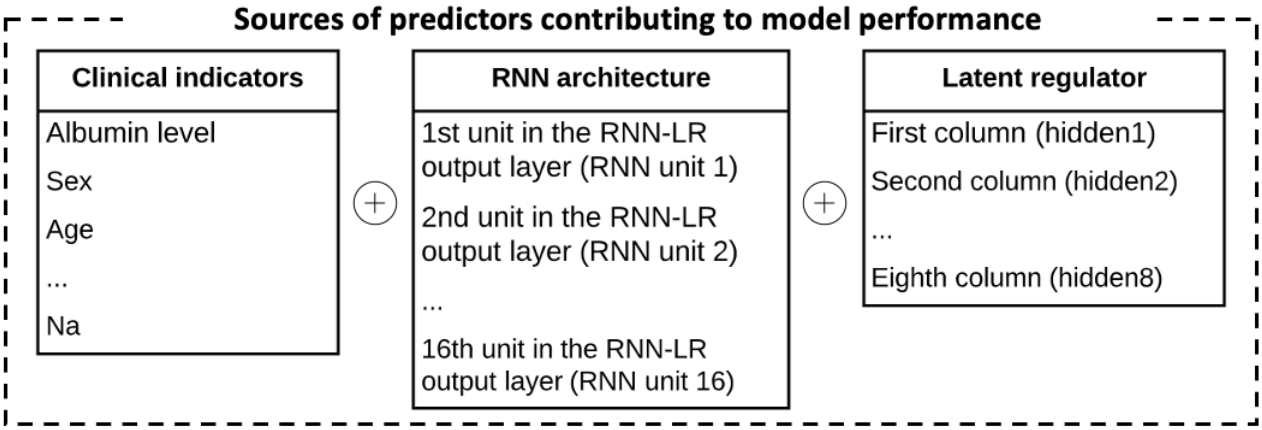
[c]HRS: hepatorenal syndrome.

[d]AUC: area under the curve.

Ablation Analysis

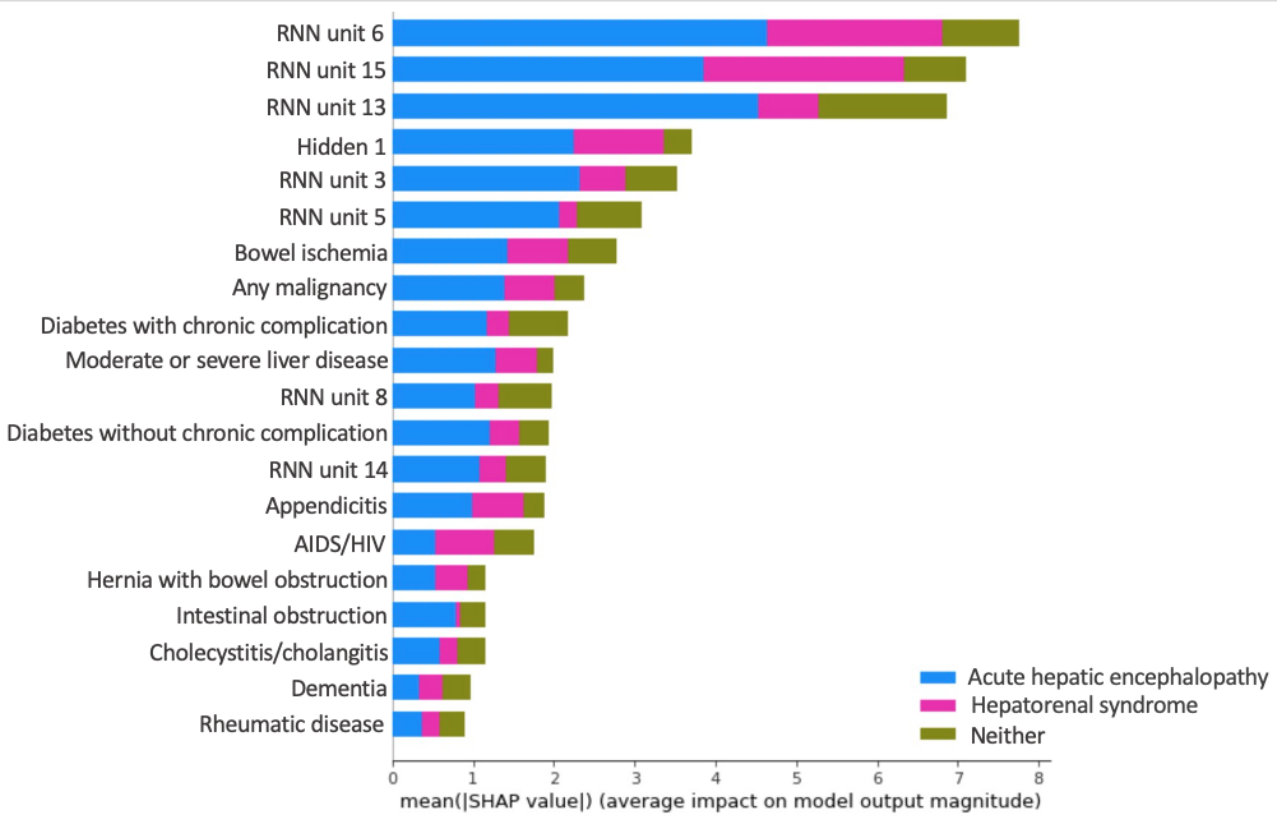
Furthermore, we analyzed the proposed deep learning–based method with deep SHAP [80] to reveal the elements that contribute more predictive power. In essence, SHAP follows a game theoretic approach to analyze the output of a predictive model and indicate the marginal contributions of different features to predictions [80]. We categorized the predictors of our RNN-LR method as clinical indicators (eg, patient's age, sex, and albumin level), RNN architecture (data representations in different units of RNN-LR output), and latent regulator (vectors in the parameter matrix), as shown in Figure 8. Having specified the different types of predictors, we used deep SHAP to identify variables that contribute more to our proposed method's ability to predict complication phenotypes accurately.

Figure 8. Different predictors in our method and their contributions to predictive power. RNN: recurrent neural network; RNN-LR: recurrent neural network-latent regulator; Na: potassium.



In Figure 9, we sort the different features according to the mean SHAP values, which are approximations of their contributions to the predictions. As shown in the figure, *RNN unit 6* contributes most to the method's predictive utilities; that is, the sixth unit of the deep learning output provides the most valuable information to predict complication phenotypes. This feature is derived from diagnostic clinical outcomes (Table 1) and conveys the value of patient representation. Overall, the SHAP values indicate that 7 of the top 20 predictors relate to the RNN architecture, that is, the method's architecture provides more important information to predict crucial complication phenotypes than clinical indicators or the latent regulator. Also, *hidden 1* refers to the first column vector of the latent regulator and is the fourth most important predictor, which confirms that the latent regulator contains important information for predicting AHE and HRS. Bowel ischemia and malignancy are two important clinical indicators for predicting crucial complication phenotypes.

Figure 9. Mean SHAP values of different predictors in the proposed recurrent neural network-latent regulator method. RNN: recurrent neural network.



Summary

The evaluation results demonstrate the advantages of the proposed method (RNN-LR) over several prevalent techniques. Our method outperforms temporal MMCBR [10] because it employs a recurrent neural network to learn the underlying features of the patient's condition and disease progression, rather than relying on available clinical data to perform clustering analyses. Time fusion CNN measures pairwise similarity in patient progression to predict complication phenotypes. In the presence of substantial missing values in patient data, as illustrated in our sample, pairwise similarity may not effectively capture clinical progression variations, which confines the predictive power of time fusion CNN [12]. Patient data, such as laboratory results, are gathered at different frequencies and time intervals. To address such temporal heterogeneities, T-LSTM learns patient representation from input (patient) data with different time intervals [11]. The proposed method instead imputes missing values in patient data to generate same-length input (data) sequences and achieves better predictive performance than T-LSTM. In particular, the use of a latent regulator, as an additional parameter matrix, to mitigate the information insufficiency constraint helps in capturing the underlying relationships of clinical factors to produce better predictions, together with the available patient clinical data. In summary, the proposed deep learning–based method addresses imbalanced outcome distributions in patient data and considers patient-level temporal heterogeneities to predict AHE and HRS by incorporating both a latent regulator and cost-sensitive analysis to extend back-propagation learning in deep neural networks.

Discussion

Principal Findings

This study offers several implications for health informatics and improved acute disease patient management. First, data insufficiencies represent a challenge to physicians' patient care and management. This study highlights the promising use of a latent parameter matrix to alleviate this constraint by demonstrating its feasibility and clinical value in the prediction of crucial complication phenotypes. This latent parameter matrix can be modified or extended to accommodate other variables or hidden risk factors to more effectively predict important patient outcomes. Second, patient data are temporally heterogeneous, which creates another difficulty for clinically using EHRs and predictive analytics. Such heterogeneities can be addressed with effective missing data imputations that learn temporal feature representations from patient data to render increased predictive utilities. Unlike many existing techniques that overlook temporal heterogeneities or inconsistencies in patient data [56,81], we illustrate that an explicit incorporation of an effective representation for temporal heterogeneities can improve predictive performance. Third, imbalanced distributions of patient outcomes prevail in clinical scenarios, which creates an additional difficulty for leveraging predictive analytics in health care. Although only a small proportion of patients develop severe complication phenotypes, the outcomes can be harmful or even fatal. We demonstrate the value of cost-sensitive learning for an increased efficacy in crucial phenotype predictions (AHE or HRS). Effective patient representation, such as short-term temporal representations from limited observed patient data, and a latent regulator jointly enable

patient information abstraction at multiple levels to predict complication phenotypes more accurately.

Our research also has important implications for clinical practice. Health care is going through a paradigmatic shift from reactive care to preventive care. Predicting important clinical events and patient outcomes, especially among patients with acute diseases, is critical to the quality of care and cost containment in patient management. The proposed method can be applied to develop advanced clinical decision support systems that assist physicians at the point of care. For example, a timely detection of patients who are likely to develop severe complications is critical but challenging. Through support by decision support systems enabled by the proposed method, physicians can identify at-risk patients and perform thorough monitoring and timely interventions to improve those patients' outcomes and well-being. Our method can also benefit health care organizations in their resource planning and allocation. For example, effective phenotype predictions can help a hospital distinguish patients who are likely or not likely to develop serious complications, so their readmission risk or length of stay can be reduced. Such benefits have important implications for resource utilization efficiency and cost containment in patient care and management.

Conclusions and Future Research

We have developed a deep learning–based method to predict crucial complication phenotypes of an acute disease. Furthermore, we have evaluated the proposed method and several prevalent benchmark techniques with a peritonitis data set by comparing their predictions of AHE and HRS. The empirical results reveal the advantageous predictive power of our method, which can address challenges pertaining to data insufficiency, temporal heterogeneity, and imbalanced outcome distributions. This study makes several contributions to the predictive analytics for an improved care and management of patients with acute diseases. First, we demonstrate the feasibility and clinical value of using a latent regulator to cope with insufficiencies in available patient data to improve phenotype predictions. The latent regulator, incorporated in the proposed method, can be expanded to model other external variables and hidden risk factors for predicting different complication phenotypes. Second, our proposed method incorporates missing data imputation and addresses temporal heterogeneities that exist in patient data, a fundamental challenge in using EHRs and predictive analytics for patient care and management. As we illustrate, temporal feature representation can be learned from patient data to provide increased predictive utilities. Third, imbalanced data prevail in clinical scenarios. Although only a relatively small proportion of patients develop severe complication phenotypes, the outcomes can be fatal. Toward that end, the proposed method reveals the value of cost-sensitive learning to address the data imbalance issue and demonstrates greater effectiveness to predict the minority class (eg, AHE and HRS), which is clinically important.

This study has several limitations that warrant continued research attention. First, the sample for the evaluation was relatively limited in size with respect to the disease category. Continued research should re-examine the proposed method with additional, diverse patient samples and different acute diseases. Second, we rely on domain experts to guide clinical feature extraction in this study. We acknowledge that some potentially important factors might be overlooked by domain experts. In addition, other complications may involve more complex risk factors, such as patient comorbidity, disease progression, and environmental factors. Thus, further research should explore how representation learning might identify features automatically from various patient clinical and behavioral data. Third, a predictive model's ability to generate interpretable results is desirable and important; however, interpreting the proposed method's predictions is difficult because its deep learning model maps input variables (eg, laboratory results, sex, age) to a numerical output variable through multiple layers. The complex structures make its prediction results difficult to interpret, unlike rule- or inductive decision tree–based techniques that can reveal interpretable relationships between input variables and the target variable. Ongoing research should explore *interpretable* computational methods built on explainable artificial intelligence. In a related sense, our method uses a latent regulator to account for observed disease progression and underlying mechanisms (eg, hidden disease patterns), so its processing and results cannot explain the underlying causes of the phenotypes. Continued efforts are needed to specify and test probable mechanisms and pathogeneses leading to crucial hepatic complications, as manifested by these phenotypes.

Acknowledgments

The authors thank the Research Institute of Chang Gung Memorial Hospital for their assistance in data collection and preprocessing. The study was partially funded by the Chang Gung Memorial Hospital Research Projects CLRPG2J0011, CMRPG2H0342, and CRRPG2H0062.

Conflicts of Interest

None declared.

References

1. Rimola A, García-Tsao G, Navasa M, Piddock LJ, Planas R, Bernard B, et al. Diagnosis, treatment and prophylaxis of spontaneous bacterial peritonitis: a consensus document. International Ascites Club. J Hepatol 2000 Jan;32(1):142-153. [doi: 10.1016/s0168-8278(00)80201-9] [Medline: 10673079]

2. Acute (medicine). Wikipedia. 2017. URL: https://en.wikipedia.org/wiki/Acute_(medicine) [accessed 2021-04-02]
3. Scheuermann RH, Ceusters W, Smith B. Toward an ontological treatment of disease and diagnosis. Summit Transl Bioinform 2009 Mar 01;2009:116-120 [FREE Full text] [Medline: 21347182]
4. Miravitlles M, Calle M, Soler-Cataluña JJ. Clinical phenotypes of COPD: identification, definition and implications for guidelines. Arch Bronconeumol 2012 Mar;48(3):86-98. [doi: 10.1016/j.arbres.2011.10.007] [Medline: 22196477]
5. Basile AO, Ritchie MD. Informatics and machine learning to define the phenotype. Expert Rev Mol Diagn 2018 Mar;18(3):219-226 [FREE Full text] [doi: 10.1080/14737159.2018.1439380] [Medline: 29431517]
6. Srinivasan A, Kaiser K. Information deficiency: implications for information systems design. In: Computer Science. Sutton: ICIS; 1981.
7. Blei AT, Córdoba J, Practice Parameters Committee of the American College of Gastroenterology. Hepatic encephalopathy. Am J Gastroenterol 2001 Jul;96(7):1968-1976. [doi: 10.1111/j.1572-0241.2001.03964.x] [Medline: 11467622]
8. Ginès P, Solà E, Angeli P, Wong F, Nadim MK, Kamath PS. Author correction: hepatorenal syndrome. Nat Rev Dis Primers 2018 Oct 15;4(1):33. [doi: 10.1038/s41572-018-0035-2] [Medline: 30323303]
9. Vaquero J, Polson J, Chung C, Helenowski I, Schiodt FV, Reisch J, et al. Infection and the progression of hepatic encephalopathy in acute liver failure. Gastroenterology 2003 Sep;125(3):755-764. [doi: 10.1016/s0016-5085(03)01051-5] [Medline: 12949721]
10. Ping X, Tseng Y, Lin Y, Chiu H, Lai F, Liang J, et al. A multiple measurements case-based reasoning method for predicting recurrent status of liver cancer patients. Computers in Industry 2015 May;69:12-21. [doi: 10.1016/j.compind.2015.01.007]
11. Baytas I, Xiao C, Zhang X, Wang F, Jain A, Zhou J, editors. Patient subtyping via time-aware LSTM networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery; 2017 Aug Presented at: KDD '17: The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August, 2017; Halifax NS Canada p. 65-74. [doi: 10.1145/3097983.3097997]
12. Suo Q, Ma F, Yuan Y, Huai M, Zhong W, Zhang A. Personalized disease prediction using a CNN-based similarity learning method. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Piscataway, New Jersey, United States: IEEE; 2017 Presented at: Personalized Disease Prediction Using a CNN-Based Similarity Learning Method. Bioinformatics and Biomedicine (BIBM), IEEE International Conference on; 13-16 Nov. 2017; Kansas City, MO, USA URL: https://ieeexplore.ieee.org/document/8217759 [doi: 10.1109/bibm.2017.8217759]
13. Carja O, Plotkin JB. The evolutionary advantage of heritable phenotypic heterogeneity. Sci Rep 2017 Jul 11;7(1):5090 [FREE Full text] [doi: 10.1038/s41598-017-05214-2] [Medline: 28698577]
14. Papa G, Degano C, Iurato MP, Licciardello C, Maiorana R, Finocchiaro C. Macrovascular complication phenotypes in type 2 diabetic patients. Cardiovasc Diabetol 2013 Jan 18;12:20 [FREE Full text] [doi: 10.1186/1475-2840-12-20] [Medline: 23331854]
15. Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. J Am Med Inform Assoc 2013 Dec;20(e2):206-211 [FREE Full text] [doi: 10.1136/amiajnl-2013-002428] [Medline: 24302669]
16. Such J, Runyon BA. Spontaneous bacterial peritonitis. Clin Infect Dis 1998 Oct;27(4):669-675. [doi: 10.1086/514940] [Medline: 9798013]
17. Gallaher JR, Cairns B, Varela C, Charles AG. Mortality after peritonitis in sub-Saharan Africa: an issue of access to care. JAMA Surg 2017 Apr 01;152(4):408-410. [doi: 10.1001/jamasurg.2016.4638] [Medline: 28030662]
18. Fernandez J, Navasa M, Planas R, Montoliu S, Monfort D, Soriano G, et al. Primary prophylaxis of spontaneous bacterial peritonitis delays hepatorenal syndrome and improves survival in cirrhosis. Gastroenterology 2007 Sep;133(3):818-824. [doi: 10.1053/j.gastro.2007.06.065] [Medline: 17854593]
19. Piotrowski D, Boroń-Kaczmarska A. Bacterial infections and hepatic encephalopathy in liver cirrhosis-prophylaxis and treatment. Adv Med Sci 2017 Sep;62(2):345-356. [doi: 10.1016/j.advms.2016.11.009] [Medline: 28514703]
20. Ng CKF, Chan MHM, Tai MHL, Lam CWK. Hepatorenal syndrome. Clin Biochem Rev 2007 Feb;28(1):11-17 [FREE Full text] [Medline: 17603637]
21. Venkataraman S, Peter S, Chandy G. Assessing the severity of hepatic encephalopathy. Indian J Gastroenterol 2003 Dec;22 Suppl 2:7-10. [Medline: 15025245]
22. Mullen K. Review of the final report of the 1998 Working Party on definition, nomenclature and diagnosis of hepatic encephalopathy. Aliment Pharmacol Ther 2007 Feb;25 Suppl 1:11-16 [FREE Full text] [doi: 10.1111/j.1746-6342.2006.03216.x] [Medline: 17295847]
23. Riordan SM, Williams R. Treatment of hepatic encephalopathy. N Engl J Med 1997 Aug 14;337(7):473-479. [doi: 10.1056/nejm199708143370707]
24. Jepsen P, Ott P, Andersen PK, Sørensen HT, Vilstrup H. Clinical course of alcoholic liver cirrhosis: a Danish population-based cohort study. Hepatology 2010 May;51(5):1675-1682. [doi: 10.1002/hep.23500] [Medline: 20186844]
25. Andrew SE, Goldberg YP, Kremer B, Telenius H, Theilmann J, Adam S, et al. The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. Nat Genet 1993 Aug;4(4):398-403. [doi: 10.1038/ng0893-398] [Medline: 8401589]

26. Alward WL, Fingert JH, Coote MA, Johnson AT, Lerner SF, Junqua D, et al. Clinical features associated with mutations in the chromosome 1 open-angle glaucoma gene (GLC1A). N Engl J Med 1998 Apr 09;338(15):1022-1027. [doi: 10.1056/NEJM199804093381503] [Medline: 9535666]
27. Overby CL, Pathak J, Gottesman O, Haerian K, Perotte A, Murphy S, et al. A collaborative approach to developing an electronic health record phenotyping algorithm for drug-induced liver injury. J Am Med Inform Assoc 2013 Dec;20(e2):243-252 [FREE Full text] [doi: 10.1136/amiajnl-2013-001930] [Medline: 23837993]
28. Richesson R, Smerek M. Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials: Duke University. Duke University. 2014. URL: https://sites.duke.edu/rethinkingclinicaltrials/ [accessed 2021-01-22]
29. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, eMERGE Network. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 2013 Oct;15(10):761-771 [FREE Full text] [doi: 10.1038/gim.2013.72] [Medline: 23743551]
30. Zhang J, Friberg IM, Kift-Morgan A, Parekh G, Morgan MP, Liuzzi AR, et al. Machine-learning algorithms define pathogen-specific local immune fingerprints in peritoneal dialysis patients with bacterial infections. Kidney Int 2017 Jul;92(1):179-191 [FREE Full text] [doi: 10.1016/j.kint.2017.01.017] [Medline: 28318629]
31. Newby C, Heaney LG, Menzies-Gow A, Niven RM, Mansur A, Bucknall C, British Thoracic Society Severe Refractory Asthma Network. Statistical cluster analysis of the British Thoracic Society Severe refractory Asthma Registry: clinical outcomes and phenotype stability. PLoS One 2014;9(7):e102987 [FREE Full text] [doi: 10.1371/journal.pone.0102987] [Medline: 25058007]
32. Liu C, Wang F, Hu J, Xiong H. Temporal phenotyping from longitudinal electronic health records: a graph based framework. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery; 2015 Aug Presented at: KDD '15: The 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August, 2015; Sydney NSW Australia p. 705-714. [doi: 10.1145/2783258.2783352]
33. Chen L, Li X, Sheng Q, Peng W, Bennett J. Mining health examination records—a graph-based approach. IEEE Transactions on Knowledge and Data Engineering 2016 Sep;28(9):2423-2437 [FREE Full text]
34. Xu X, Cui L, Liu S, Li H, Liu L, Zheng Y. Predicting hospital readmission from longitudinal healthcare data using graph pattern mining based temporal phenotypes. 2017 Presented at: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); 13-16 Nov. 2017; Kansas City, MO, USA. [doi: 10.1109/bibm.2017.8217761]
35. Pivovarov R, Perotte AJ, Grave E, Angiolillo J, Wiggins CH, Elhadad N. Learning probabilistic phenotypes from heterogeneous EHR data. J Biomed Inform 2015 Dec;58:156-165 [FREE Full text] [doi: 10.1016/j.jbi.2015.10.001] [Medline: 26464024]
36. Ross JC, Castaldi PJ, Cho MH, Chen J, Chang Y, Dy JG, et al. A bayesian nonparametric model for disease subtyping: application to emphysema phenotypes. IEEE Trans Med Imaging 2017 Jan;36(1):343-354. [doi: 10.1109/tmi.2016.2608782]
37. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018 Nov 27;19(6):1236-1246 [FREE Full text] [doi: 10.1093/bib/bbx044] [Medline: 28481991]
38. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
39. Ho J, Ghosh J, Sun J, editors. Marble: high-throughput phenotyping from electronic health records via sparse nonnegative tensor factorization. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: Association for Computing Machinery; 2014 Aug Presented at: KDD '14: The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August, 2014; New York USA p. 115-124. [doi: 10.1145/2623330.2623658]
40. Kale D, Che Z, Bahadori M, Li W, Liu Y, Wetzel R. Causal phenotype discovery via deep networks. In: AMIA Annual Symposium Proceedings. 2015 Nov 5 Presented at: American Medical Informatics Association; 2015; San Francisco p. 677-686 URL: https://pubmed.ncbi.nlm.nih.gov/26958203/
41. Guo W, Xu Y, Feng X. Deepmetabolism: a deep learning system to predict phenotype from genome sequencing. arXiv 2017:A [FREE Full text] [doi: 10.1101/135574]
42. Agache I, Akdis C. Precision medicine and phenotypes, endotypes, genotypes, regiotypes, and theratypes of allergic diseases. J of Clin Investi 2019;129(4):-. [doi: 10.1172/jci124611]
43. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. J Am Med Assoc 2016 Dec 13;316(22):2402-2410. [doi: 10.1001/jama.2016.17216] [Medline: 27898976]
44. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan 7;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
45. Pouladi F, Salehinejad H, Gilani A. Recurrent neural networks for sequential phenotype prediction in genomics. In: Recurrent Neural Networks for Sequential Phenotype Prediction in Genomics. 2015 Presented at: 2015 International Conference on Developments of E-Systems Engineering (DeSE); 13-14 Dec, 2015; Duai, United Arab Emirates. [doi: 10.1109/dese.2015.52]

46. Lasko TA, Denny JC, Levy MA. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. PLoS One 2013 Jun 24;8(6):66341 [FREE Full text] [doi: 10.1371/journal.pone.0066341] [Medline: 23826094]
47. Gupta P, Malhotra P, Vig L, Shroff G. Transfer learning for clinical time series analysis using recurrent neural networks. arXiv :180701705 Preprint posted online July 4, 2018. [FREE Full text]
48. Ma F, Chitta R, Zhou J, You Q, Sun T, Gao J. Dipole: diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Dipole; 2017 Aug Presented at: KDD '17: The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August, 2017; Halifax NS Canada p. 1903-1911. [doi: 10.1145/3097983.3098088]
49. Lee C, Chen YPP. Prediction of drug adverse events using deep learning in pharmaceutical discovery. Brief Bioinform 2020 Apr 29:2020. [doi: 10.1093/bib/bbaa040] [Medline: 32349125]
50. Tan Q, Ma A, Ye M, Yang B, Deng H, Wong V. UA-CRNN: Uncertainty-aware convolutional recurrent neural network for mortality risk prediction. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York: Association for Computing Machinery; 2019 Nov Presented at: CIKM '19: The 28th ACM International Conference on Information and Knowledge Management; November, 2019; Beijing China p. 109-118. [doi: 10.1145/3357384.3357884]
51. Cao P, Zhao D, Zaiane O. An optimized cost-sensitive SVM for imbalanced data learning. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. 2013 Presented at: PAKDD 2013: Advances in Knowledge Discovery and Data Mining; 2013; old Coast, Australia p. 280-292. [doi: 10.1007/978-3-642-37456-2_24]
52. Thai-Nghe N, Gantner Z, Schmidt-Thieme L. Cost-sensitive learning methods for imbalanced data. In: The 2010 International Joint Conference on Neural Networks (IJCNN). 2010 Presented at: The 2010 International Joint Conference on Neural Networks (IJCNN); July 18-23, 2010; Barcelona, Spain. [doi: 10.1109/ijcnn.2010.5596486]
53. Kolb H, Schneider B, Heinemann L, Heise T, Lodwig V, Tshiananga JKT, et al. Type 2 diabetes phenotype and progression is significantly different if diagnosed before versus after 65 years of age. J Diabetes Sci Technol 2008 Jan;2(1):82-90 [FREE Full text] [doi: 10.1177/193229680800200112] [Medline: 19885181]
54. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, et al. Identification of asthma phenotypes using cluster analysis in the severe asthma research program. Am J Respir Crit Care Med 2010 Feb 15;181(4):315-323. [doi: 10.1164/rccm.200906-0896oc]
55. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, et al. Cluster analysis and clinical asthma phenotypes. Am J Respir Crit Care Med 2008 Aug;178(3):218-224. [doi: 10.1164/rccm.200711-1754oc]
56. Royall DR, Salazar R, Palmer RF. Latent variables may be useful in pain's assessment. Health Qual Life Outcomes 2014;12(1):13. [doi: 10.1186/1477-7525-12-13]
57. Sterne J, White I, Carlin J, Spratt M, Royston P, Kenward M, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009 Jun 29;338:2393 [FREE Full text] [doi: 10.1136/bmj.b2393] [Medline: 19564179]
58. Xu D, Hu PJ, Huang T, Fang X, Hsu C. A deep learning-based, unsupervised method to impute missing values in electronic health records for improved patient management. J Biomed Inform 2020 Nov;111:103576. [doi: 10.1016/j.jbi.2020.103576] [Medline: 33010424]
59. Zhang H, Xie P, Xing E. Missing value imputation based on deep generative models. Computer Science, Machine Learning 2018:- [FREE Full text]
60. Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor. Newsl 2004 Jun;6(1):20-29. [doi: 10.1145/1007730.1007735]
61. Mora A, Krug BS, Grigonis AM, Dawson A, Jing Y, Hammerman SI. Optimizing laboratory test utilization in long-term acute care hospitals. Proc (Bayl Univ Med Cent) 2017 Jan 11;30(1):26-29 [FREE Full text] [doi: 10.1080/08998280.2017.11929516] [Medline: 28127124]
62. National Health Insurance Research Database, Taiwan. 2014. URL: https://nhird.nhri.org.tw/en/ [accessed 2021-01-22]
63. Hsing AW, Ioannidis JPA. Nationwide Population Science: lessons from the Taiwan National Health Insurance Research Database. JAMA Intern Med 2015 Sep 01;175(9):1527-1529. [doi: 10.1001/jamainternmed.2015.3540] [Medline: 26192815]
64. Shao S, Chan Y, Kao Yang Y, Lin S, Hung M, Chien R, et al. The Chang Gung Research Database-A multi-institutional electronic medical records database for real-world epidemiological studies in Taiwan. Pharmacoepidemiol Drug Saf 2019 May 16;28(5):593-600. [doi: 10.1002/pds.4713] [Medline: 30648314]
65. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005 Nov;43(11):1130-1139. [doi: 10.1097/01.mlr.0000182534.19832.83] [Medline: 16224307]
66. Huang T, Huang S, Shyu Y, Lee C, Jwo S, Chen P, et al. A procalcitonin-based algorithm to guide antibiotic therapy in secondary peritonitis following emergency surgery: a prospective study with propensity score matching analysis. PLoS One 2014 Mar 4;9(3):90539 [FREE Full text] [doi: 10.1371/journal.pone.0090539] [Medline: 24594916]
67. Fletcher J. What to know about hepatic encephalopathy. Medical News Today. 2019 Nov 6. URL: https://www.medicalnewstoday.com/articles/326919 [accessed 2021-01-11]

68. Facciorusso A. Hepatorenal syndrome Type 1: current challenges and future prospects. Ther Clin Risk Manag 2019 Nov;15:1383-1391. [doi: 10.2147/tcrm.s205328]
69. Fujiwara K, Mochida S. Indications and criteria for liver transplantation for fulminant hepatic failure. J Gastroenterol 2002 Jan;37(S13):74-77. [doi: 10.1007/bf02990104]
70. Omagari K, Norimura D, Yajima H, Kadokawa Y, Takamori K, Yoshimura M. Recovery from fulminant hepatic failure and prolonged deep coma without liver transplantation. Acta Medica Nagasakiensia 2004;49(4):153-157 [FREE Full text]
71. Salerno F, Gerbes A, Ginès P, Wong F, Arroyo V. Diagnosis, prevention and treatment of hepatorenal syndrome in cirrhosis. Postgrad Med J 2008 Dec 01;84(998):662-670. [doi: 10.1136/gut.2006.107789] [Medline: 19201943]
72. Inadomi J, Sonnenberg A. Cost-analysis of prophylactic antibiotics in spontaneous bacterial peritonitis. Gastroenterology 1997 Oct;113(4):1289-1294. [doi: 10.1053/gast.1997.v113.pm9322524] [Medline: 9322524]
73. Domhan T, Springenberg J, Hutter F. Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves. In: University of Freiburg. 2015 Presented at: Twenty-Fourth International Joint Conference on Artificial Intelligence; 2015; Buenos Aires, Argentina URL: https://ml.informatik.uni-freiburg.de/papers/15-IJCAI-Extrapolation_of_Learning_Curves.pdf
74. Liu Y, Zheng Y. One-against-all multi-class SVM classification using reliability measures. In: Proceedings 2005 IEEE International Joint Conference on Neural Networks. 2005 Presented at: 2005 IEEE International Joint Conference on Neural Networks, 2005; 31 July-4 Aug, 2005; Montreal, Que, Canada. [doi: 10.1109/ijcnn.2005.1555963]
75. Milgram J, Cheriet M, Sabourin R. 'One Against One' or 'One Against All': which one is better for handwriting recognition with SVMs? HAL-Inria. 2006. URL: https://hal.inria.fr/inria-00103955 [accessed 2021-01-22]
76. David MW. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv 2010:- [FREE Full text] [doi: 10.9735/2229-3981]
77. van Buuren S, Groothuis-Oudshoorn K. Multivariate imputation by chained equations in R. J Stat Soft 2011;45(3):1-68. [doi: 10.18637/jss.v045.i03]
78. Mazumder R, Hastie T, Tibshirani R. Spectral regularization algorithms for learning large incomplete matrices. J Mach Learn Res 2010 Mar 01;11:2287-2322 [FREE Full text] [Medline: 21552465]
79. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, et al. Missing value estimation methods for DNA microarrays. Bioinformatics 2001 Jun 01;17(6):520-525. [doi: 10.1093/bioinformatics/17.6.520] [Medline: 11395428]
80. Lundberg S. A unified approach to interpreting model predictions. arXiv Preprint posted online November 25, 2017. [FREE Full text]
81. Hoshida Y, Villanueva A, Llovet JM. Molecular profiling to predict hepatocellular carcinoma outcome. Expert Rev Gastroenterol Hepatol 2009 Apr;3(2):101-103 [FREE Full text] [doi: 10.1586/egh.09.5] [Medline: 19351279]

Abbreviations

AHE: acute hepatic encephalopathy
AUC: area under the curve
CGRD: Chang Gung Research Database
CNN: convolutional neural network
EHR: electronic health record
HRS: hepatorenal syndrome
ICU: intensive care unit
MMCBR: multiple measurement case-based reasoning
NHIRD: National Health Insurance Research Database
RNN: recurrent neural network
RNN-LR: recurrent neural network-latent regulator
SHAP: Shapley Additive Explanations
SMOTEENN: Synthetic Minority Oversampling Technique–Edited Nearest Neighbors
T-LSTM: time aware long short-term memory

Edited by G Eysenbach; submitted 22.02.20; peer-reviewed by OS Liang, A Aminbeidokhti, J Yang, N Mohammad Gholi Mezerji; comments to author 29.06.20; revised version received 13.09.20; accepted 21.12.20; published 12.02.21.

Please cite as:
Sheng JQ, Hu PJH, Liu X, Huang TS, Chen YH
Predictive Analytics for Care and Management of Patients With Acute Diseases: Deep Learning–Based Method to Predict Crucial Complication Phenotypes
J Med Internet Res 2021;23(2):e18372
URL: http://www.jmir.org/2021/2/e18372/
doi:10.2196/18372
PMID:33576744

Original Paper

Classification Accuracy of Hepatitis C Virus Infection Outcome: Data Mining Approach

Mario Frias[1*], PhD; Jose M Moyano[2,3*], PhD; Antonio Rivero-Juarez[1], PhD; Jose M Luna[2,3], PhD; Ángela Camacho[1], MD, PhD; Habib M Fardoun[4], PhD; Isabel Machuca[1], MD, PhD; Mohamed Al-Twijri[4], MSc; Antonio Rivero[1], MD, PhD; Sebastian Ventura[2,3], PhD

[1]Department of Clinical Virology and Zoonoses, Maimonides Biomedical Research Institute of Córdoba, Córdoba, Spain

[2]Department of Computer Science and Numerical Analysis, University of Córdoba, Córdoba, Spain

[3]Knowledge Discovery and Intelligent Systems in Biomedicine Laboratory, Maimonides Biomedical Research Institute of Córdoba, Córdoba, Spain

[4]Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

[*]these authors contributed equally

Corresponding Author:
Antonio Rivero, MD, PhD
Department of Clinical Virology and Zoonoses
Maimonides Biomedical Research Institute of Córdoba
Avenida Menéndez Pidal s/n
Córdoba, 14004
Spain
Phone: 34 957213806
Email: ariveror@gmail.com

Abstract

Background: The dataset from genes used to predict hepatitis C virus outcome was evaluated in a previous study using a conventional statistical methodology.

Objective: The aim of this study was to reanalyze this same dataset using the data mining approach in order to find models that improve the classification accuracy of the genes studied.

Methods: We built predictive models using different subsets of factors, selected according to their importance in predicting patient classification. We then evaluated each independent model and also a combination of them, leading to a better predictive model.

Results: Our data mining approach identified genetic patterns that escaped detection using conventional statistics. More specifically, the partial decision trees and ensemble models increased the classification accuracy of hepatitis C virus outcome compared with conventional methods.

Conclusions: Data mining can be used more extensively in biomedicine, facilitating knowledge building and management of human diseases.

(J Med Internet Res 2021;23(2):e18766) doi:10.2196/18766

KEYWORDS

HIV/HCV; data mining; PART; ensemble; classification accuracy

Introduction

Univariate and multivariate analysis are the two main conventional approaches to statistical analysis in the scientific method. Multivariate analysis in particular is used to determine the contribution of several factors (risk factors in biomedicine) to a single event or result. Genome-wide association studies (GWAS) have been widely used in case-control settings to identify which genetic variants, known as single nucleotide polymorphisms (SNPs), are associated with human diseases or traits [1,2]. In biomedicine, a number of studies have performed univariate and multivariate analyses based on the results of GWAS in order to obtain new risk or protective factors.

The 2017 study by our group using this method analyzed two groups of patients diagnosed with hepatitis C virus (HCV) infection [3]. One group consisted of patients who experienced

spontaneous resolution of infection during first 6 months of infection (acute phase) and the other of patients who developed chronic hepatitis C. It is important from a clinical point of view to have tools available to predict HCV outcome (whether spontaneous resolution or chronic hepatitis C). With this in mind, one GWAS identified an SNP in the interferon lambda–3 (IFNL3) gene as a factor in spontaneous resolution [1]. That study showed that patients with the CC IFNL3 genotype had a greater likelihood of experiencing spontaneous resolution, while patients with the non-CC IFNL3 genotype were more likely to develop chronic hepatitis C. In our previous study, we studied whether haplotypes of the human leukocyte antigen (HLA) and killer cell immunoglobulin-like receptor (KIR) improved the predictive capacity of the IFNL3 genotype and found that different combinations of these genes (HLA-B44, HLA-C12, and KIR3DS1), together with the IFNL3 genotype, increased the classification accuracy of HCV outcome. More specifically, based on this combination of genes, a patient could be classified as having a genetically unfavorable profile (GUP) or a genetically favorable profile (GFP) for spontaneous resolution of HCV infection.

Data mining—the process of extracting hidden associations in datasets—is a promising trend in biomedicine and important for identifying factors that are never discovered by conventional statistical methods [4-6]. A number of studies have demonstrated the effectiveness of data mining techniques in biomedicine. Examples include the application of feature selection methods to reduce the dimensionality of biomedical problems [7], identification of key genes to improve the accuracy of classification models [8], and the use of big data techniques in scenarios where there is a large volume of data [9]. The aim of our study therefore was to reanalyze the dataset used in our 2017 study [3] using data mining approaches in order to find models that improved the classification accuracy of the genes studied.

Methods

Dataset Description and Data Preprocessing

This study was completed using a dataset from our earlier research [3], which was performed between 2013 and 2017 on 138 individuals, all of whom were HIV/HCV coinfected patients from the infectious diseases unit at the Hospital Reina Sofía in Cordoba (Spain). The patients were categorized as chronic hepatitis C or spontaneous resolution. Patients with spontaneous resolution were those who had undetectable HCV viral loads during the acute phase of infection and did not require specific treatment; patients with chronic hepatitis C were those who had detectable HCV viral loads after the acute phase and needed treatment to be cured. Further information about spontaneous resolution and chronic hepatitis C and an analysis of the IFNL3, KIR, HLA-B, and HLA-C genes is published elsewhere [3].

The dataset comprises 43 input features from different markers in every patient. The markers were IFNL3 genotype (1 feature), HLA-B (17 features), epitope Bw (1 feature), HLA-C (12 feature), and KIR genotype (12 features). The input features of each marker are shown in Table 1. To prevent great loss of information, the data from patients with missing values in any of the input features were completed using the k-nearest neighbors imputation method (k=3) before conducting the computational study [10]. This method finds the nearest neighbors to instances with missing values and fills in the gaps with the most frequent value in the nearest neighbors. A total of 46 of 138 patients included in this study had missing values in the features. The dataset is publicly available [11].

Table 1. Features of each variable.

IFNL3[a]	Epitope Bw[b]	HLA-B[c]	HLA-C	KIR[d]
CC[e]	Bw4	B*07	C*01	3DL1
Non-CC	Bw6	B*08	C*02	2DL1
—	Bw4/Bw6	B*14	C*03	2DL2
—	—	B*15	C*04	2DL3
—	—	B*18	C*05	2DL5
—	—	B*27	C*06	2DS1
—	—	B*35	C*07	2DS2
—	—	B*38	C*08	2DS3
—	—	B*39	C*12	2DS5
—	—	B*40	C*15	3DL2
—	—	B*44	C*16	2DP1
—	—	B*45	C*18	3DS1
—	—	B*49	—	—
—	—	B*50	—	—
—	—	B*51	—	—
—	—	B*52	—	—
—	—	B*57	—	—

[a]IFNL3: interferon lambda–3.

[b]Bw: Epitope Bw.

[c]HLA-B: human leukocyte antigen–B.

[d]KIR: killer cell immunoglobulin-like receptor.

[e]CC: Genotype CC.

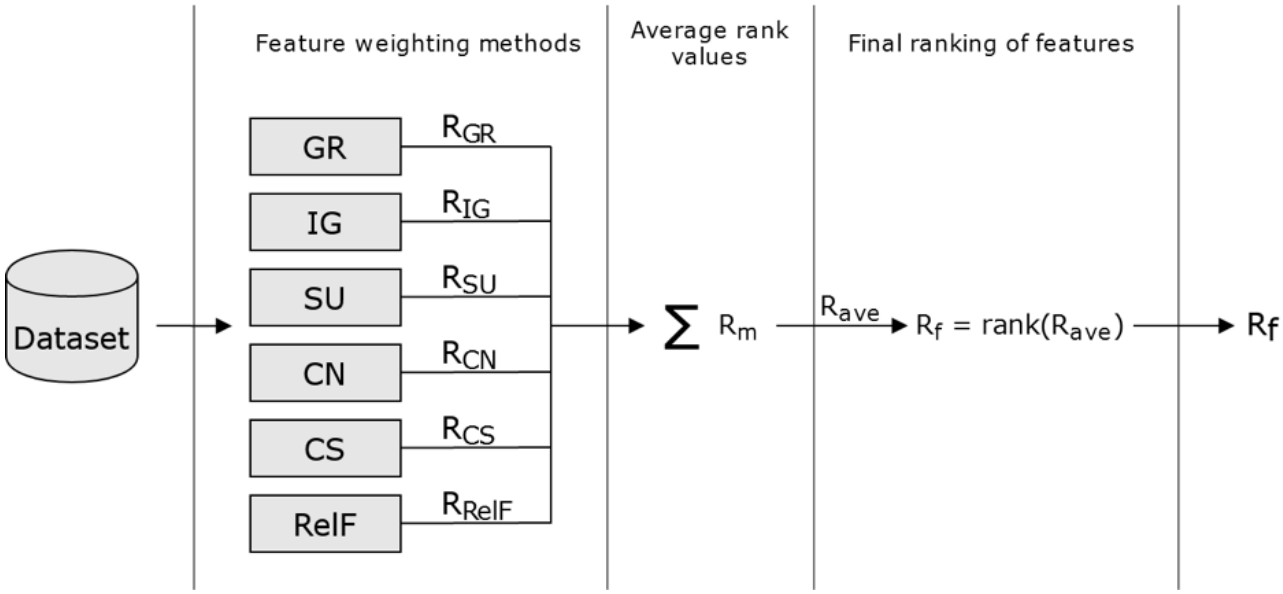
Determining the Best Subset of Features

To construct the models, we followed a procedure that has been effectively applied in recent studies [12] for selecting the subset of factors or input features that best describes the patient [13,14]. We first ranked the features in terms of relevance from highest to lowest, and then selected the best subset of features. The relevance of a feature was weighted according to its ability to distinguish the classes [15].

To avoid bias in the process of estimating feature relevance, 6 well-known feature estimation methods were used: gain ratio [16], information gain [17], symmetrical uncertainty [18], consistency [19], chi-squared [20], and relief-F [21]. These feature estimation algorithms are all supervised learning methods that use a priori classification to estimate the relevance of features but do not depend on the effectiveness of a classifier, so the biases of the learning algorithms do not influence the feature selection process. The filter methods evaluate the usefulness of a feature (or set of features) based on measures of distance, dependency, information, or correlation with data [15].

To assess the feature weighting methods and better estimate the relevance of input features, the 10-fold cross-validation procedure was executed 3 times, and the results were averaged. The relief-F method was executed with parameter values set at between 5 and 10 [13,22]. The final ranking of features was computed as follows: (1) each feature weighting method provided its own ranking of methods, R_m, with m being each individual method; (2) for each input feature, an average weight was computed as the average value of the rankings provided; (3) a final ranking, R_f, was computed given the averaged weight values, in which the feature with the highest average weight was the most important. A flowchart of the process followed in this study to compute the final ranking is shown in Figure 1. This procedure was implemented using the open source WEKA (Waikato Environment for Knowledge Analysis) data mining software (v 3.9.3) [23].

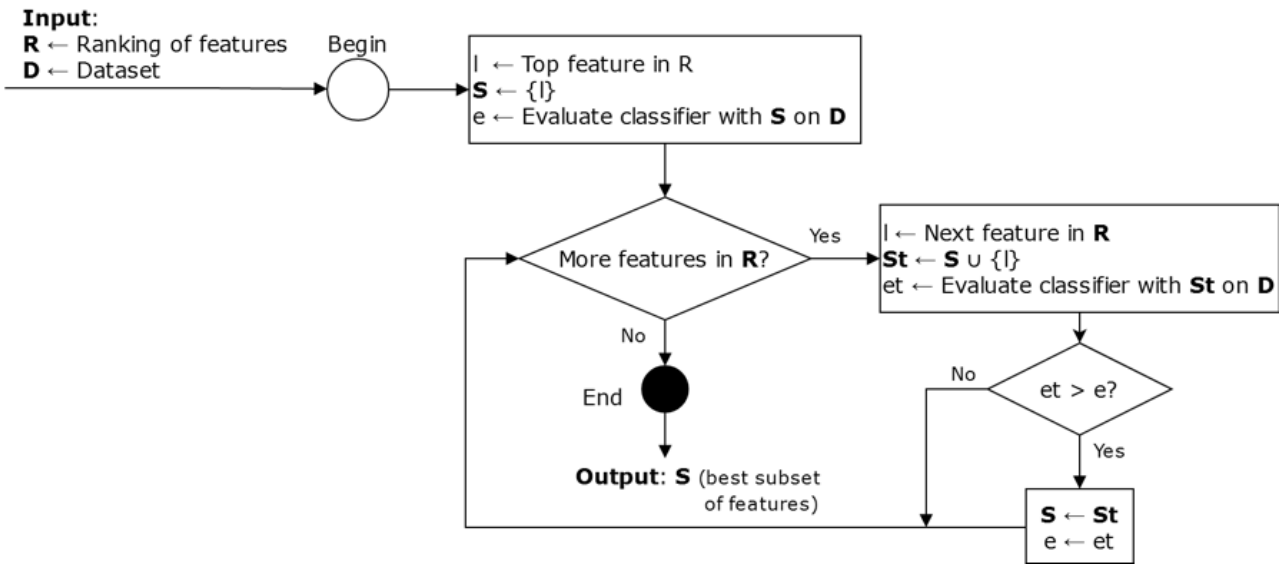
Figure 1. Flowchart illustrating the process of computing the final ranking of features.



Once the ranking of features was determined, it was then used to compute the subset of features that best predicted the class of a patient when constructing each model. We proposed a computational method that determined the best subset of features without evaluating all possible combinations of features. The combination formula used was 2^n–1 (exponential size), with n being the number of features in the dataset. This method required evaluation of a maximum of [redacted] combinations of features.

Given the final ranking of features R_f, R_f^l represented the subranking of features in which l was the highest ranked feature; in other words, the subranking of features was formed of l and all subsequent features in R_f. Figure 2 shows a flowchart of steps performed for each R_f^l subranking. Finally, each subset of features was evaluated according to the performance of the corresponding classifier in predicting whether or not a patient belonged to each group (chronic hepatitis C or spontaneous resolution).

Figure 2. Flowchart to create classifiers with optimal subsets of features.



Model Construction

The procedure to select the best ranked subset of features was performed for each classifier used. We proposed the use of several classifiers, namely, partial decision trees (PART) [24], random forest [25], sparse linear discriminant analysis [26], support vector machines with both linear and Gaussian kernels [27], and L1-regularized logistic regression [28]. The aim of the classifiers is to learn from the input features and the outcome, for any new patient, is a prediction of whether the patient is categorized as chronic hepatitis C or spontaneous resolution. The random forest, sparse linear discriminant analysis, support vector machines, and L1-regularized logistic regression methods are regarded as noninterpretable (black-box) models; they usually perform better than white-box models and the outcome given by the model is always interpretable, even though it is not easy to understand the steps taken by the model to create its outcome or prediction. PART, on the other hand, is a white-box, rule-based classifier, which is more interesting from the point of view of interpretability of the model. PART

returns a set of rules, one of which will be activated for each pattern. With respect to the set of rules, the expert can easily understand why it predicts each class; hence, although the performance of the final model is slightly reduced, it can provide the expert with valuable knowledge.

For the execution of each model, the 10-fold cross-validation procedure was repeated 3 times, evaluating on the corresponding test set in each case, thus averaging the values over a total of 30 different executions. For the algorithm parameter search, a random search of parameters was completed across 30 different combinations. The models were constructed using the caret package in R version 6.0-86 (R Foundation for Statistical Computing) [29].

Selection of Best Models

To determine the effectiveness of the models constructed in the previous section, the area under the receiver operating characteristic (AUROC) curve was used. The AUROC of each model was obtained by averaging the AUROC on the corresponding test set over all 30 executions. All models with AUROC > 0.80 were stored for further analysis and study. In order to obtain a simple and interpretable model with high precision, we also highlighted and selected the PART model with the highest accuracy on the whole patient dataset for further study.

Ensemble learning is a widely used data mining task that enhances the final predictive performance of the classifier by combining the predictions of diverse simpler classifiers [30]. The use of ensemble models definitely reduces the interpretability of simpler models but usually yields much better classification performance. We proposed, therefore, not only to use the classifiers built so far but also a combination of some of them. To select members of the ensemble, we followed a similar procedure to the one used to obtain the best subsets of features. First, all single models were ordered according to classification accuracy over all patients. An ensemble was created with the first model only and then including the following model in the ensemble: if the resulting accuracy was better than before, the model remained in the ensemble, otherwise it was removed. Once all the models were tried in the ensemble, the procedure was repeated, but without using the best model. Finally, the subset of models that performed best together in the ensemble was returned.

Comparison of Conventional and Data Mining Methodologies

Once the two models were obtained (the best PART model and the ensemble), their performance was compared with the results obtained from the previous study [3]. All the values that defined classification accuracy for chronic hepatitis C and spontaneous resolution in the models obtained in this study and the model obtained in the previous study (IFNL3, HLA-B*44, HLA-C*12, and KIR3DS1) were compared. These values included correct classification rate (CCR), positive predictive value (PPV) for spontaneous resolution, negative predictive value (NPV) for chronic hepatitis C, sensitivity, specificity, and AUROC.

The methods were compared in two analyses. The first analysis was based on all patients included in the study (n=138) and compared the classification accuracy of the IFNL3 genotype (CC or non-CC) against the models obtained in this study. A second comparison was then made that included only patients who had no missing values for the IFNL3, HLA-B, HLA-C, and KIR genes (n=104). The analysis consisted of comparing the classification accuracy of the models obtained in this study and combinations of genes obtained in the previous study. This combination of genes was used to classify a patient as having a GUP or GFP.

To avoid overfitting the training data for the models proposed in this paper, the data were split as follows. The dataset was partitioned using 10-fold cross-validation and the models trained using 9 out of 10 partitions, with the remaining one being used to evaluate the models. This was repeated so that all partitions were used once as test data. In this way, the models were evaluated using data from patients that had not been used during the learning phase, thus evaluating the ability of the models to generalize to new patients. In addition, the process was repeated on 3 different occasions using different seeds to create the partitions so that the results were consistent and not biased by the partitions created each time. Despite the lack of new data compared with the previous study, this process, which is standard in data mining, enables us to better determine whether the proposed models perform better than the previous one without this leading to overfitting the available data.

Results

Models Constructed for Spontaneous Resolution Prediction

After construction of the aforementioned classifiers from optimal subsets of features, we obtained more than 500 models with an AUROC greater than 0.80. A list of the models can be found in Multimedia Appendix 1. We focused first on the PART-1 model, which was constructed with just 4 different features: IFNL3, HLA-B*44, KIR2DS1, and KIR3DS1 (Multimedia Appendix 1). This yielded the best performance in terms of accuracy across all patients and was one of the simplest classifiers using the PART method, showing a good prediction performance but also being easy to interpret by experts. This method comprises a set of interpretable rules that are evaluated in order until one of them meets the conditions for a given example, as follows:

- IFNL3 = non-CC AND KIR3DS1 = no AND KIR2DS1 = no AND HLA.B44 = no: chronic hepatitis C (25.0/6.0)
- IFNL3 = non-CC AND KIR3DS1 = yes: chronic hepatitis C (20.0)
- HLA.B44 = yes AND KIR2DS1 = no AND IFNL3 = yes: chronic hepatitis C (12.0)
- HLA.B44 = no AND KIR3DS1 = no AND KIR2DS1 = yes AND IFNL3 = CC: spontaneous resolution (13.0/2.0)
- HLA.B44 = no AND KIR3DS1 = no AND IFNL3 = CC: spontaneous resolution (36.0/12.0)
- KIR2DS1 = no: chronic hepatitis C (11.0/2.0)
- KIR3DS1 = no AND HLA.B44 = no: spontaneous resolution (7.0/2.0)
- KIR3DS1 = yes AND HLA.B44 = no: spontaneous resolution (7.0/3.0)

- KIR3DS1 = no: spontaneous resolution (5.0)
- (default rule): chronic hepatitis C (2.0)

Given a patient, the decision algorithm first checks whether the conditions of the first rule antecedent are met; if they are, the patient is classified in the given class, otherwise, it tries the following rule. For example, the model first checks whether IFNL3 is "non-CC" and whether KIR3DS1, KIR2DS1, and HLA-B*44 are all "no." If these conditions are met for the given patient, the patient is classified as chronic hepatitis C (because from the available data, 25 of the 31 patients who met these conditions were from the chronic hepatitis C class). If the patient does not meet any of these conditions, it checks whether the conditions for the second rule are met (if IFNL3 is "non-CC" and KIR3DS1 is "yes"), and so on. If none of the antecedents is satisfied, there is a default rule at the end for those patients who do not meet the conditions of any of the rules. Hence the model is highly interpretable by a clinician, since the model itself gives the reasons for its outcome.

Following the previously mentioned procedure, we also combined a subset of 5 models into an ensemble that included 3 random forest models (RF-25, RF-28, and RF-66) and 2 PART (PART-17 and PART-10). This combination of models was expected to perform better than the single PART model (and better than each of the members of the ensemble separately), although it was less interpretable than is the PART-1 method; in other words, it was more difficult to understand why this ensemble model returned each of its predictions. The ensemble model used a total of 24 features across all the base members: IFNL3, KIR2DS1, KIR2DS2, KIR2DL2, KIR3DL1, KIR3DL2, KIR3DS1, HLA-B*14, HLA-B*18, HLA-B*35, HLA-B*38, HLA-B*39, HLA-B*44, HLA-B*50, HLA-B*57, HLA-C*02, HLA-C*03, HLA-C*04, HLA-C*06, HLA-C*07, HLA-C*08, HLA-C*12, HLA-C*18, and Epitope Bw.

Comparison of Accuracy Between the Methodologies

The classification accuracy for chronic hepatitis C and spontaneous resolution between the two methodologies (previous paper vs this study) was contrasted, and the results are set out below. Table 2 shows the results comparing the performance of the proposed methods and the IFNL3 marker using all 138 patients in the database. Table 3 compares the proposed models and the GUP/GFP model proposed in the previous study [3], in this case using only those patients without missing values for HLA-B*44, HLA-C*12, and KIR3DS1. Note that the proposed models are able to make predictions for these patients, but the GUP/GFP method is not, which is indeed one of the strengths of the models proposed in this study.

Table 2. Classification accuracy of IFNL3 (previous study) and models obtained (this study). Comparison made from the analysis of 138 patients without missing values for IFNL3.

Genotype/model	CCR[a] %	PPV[b] %	NPV[c] %	Sensitivity %	Specificity %	AUROC[d]
IFNL3[e] CC[f]/non-CC	71.0	61.6	81.5	78.9	65.4	0.72
PART[g]-1	78.6	69.6	87.3	84.3	74.8	0.85
Ensemble	82.5	77.7	86.8	81.2	83.7	0.89

[a]CCR: correct classification rate.

[b]PPV: positive predictive value for spontaneous resolution.

[c]NPV: negative predictive value for chronic hepatitis C.

[d]AUROC: area under the receiver operating characteristic curve.

[e]IFNL3: interferon lambda–3.

[f]CC: genotype CC

[g]PART: partial decision trees.

Table 3. Classification accuracy of genetically unfavorable profile/genetically favorable profile (previous study) and models obtained (this study). This analysis was performed on the 104 patients without missing values for any of the genes IFNL3, HLA-B, HLA-C, and KIR.

Profile/model	CCR[a] %	PPV[b] %	NPV[c] %	Sensitivity %	Specificity %	AUROC[d]
GUP[e]/GFP[f]	76.0	64.4	84.7	76.3	75.7	0.76
PART[g]-1	80.0	69.6	87.3	84.3	74.7	0.85
Ensemble	84.8	77.7	86.8	81.2	83.7	0.89

[a]CCR: correct classification rate.

[b]PPV: positive predictive value for spontaneous resolution.

[c]NPV: negative predictive value for chronic hepatitis C.

[d]AUROC: area under the receiver operating characteristic curve.

[e]GUP: genetically unfavorable profile.

[f]GFP: genetically favorable profile.

[g]PART: partial decision trees.

In both cases, the proposed models using the data mining techniques outperformed the earlier study. The ensemble method returned a better performance than PART-1 on most evaluation metrics, although the virtue of PART-1 is that it is an interpretable model (see rules) and the clinician could obtain useful knowledge from the list of decision rules. In the case where all 138 patients were used, the ensemble raised the CCR in the model from 71.0% in the previous study to 82.5% and the AUROC from 0.72 to 0.89. In the comparison using 104 patients without missing values in their GUP features (see Table 3), the ensemble model likewise increased accuracy from 76.0% in the previous study to 84.8% and the AUROC from 0.76 to 0.89, thus consistently demonstrating a good level of performance relative to the standard approaches to this kind of data analysis.

Table 4 presents the results of all the metrics of each member in the ensemble method. The results demonstrate that combining the predictions of simpler but less accurate models leads to better overall performance.

Table 4. Classification accuracy of models included in the ensemble. Results obtained from the analysis of 138 patients without missing values for IFNL3.

Model	CCR[a] %	PPV[b] %	NPV[c] %	Sensitivity %	Specificity %	AUROC[d]
rf[e]-25	77.1	74.9	79.5	69.4	83.5	0.88
rf-28	75.7	72.9	78.6	68.4	81.8	0.88
rf-66	76.0	70.2	81.2	73.7	78.8	0.85
PART[f]-17	78.6	74.1	84.6	78.7	79.1	0.79
PART-10	78.3	72.6	85.9	81.4	76.2	0.85

[a]CCR: correct classification rate.

[b]PPV: positive predictive value for spontaneous resolution.

[c]NPV: negative predictive value for chronic hepatitis C.

[d]AUROC: area under the receiver operating characteristic curve.

[e]rf: random forest.

[f]PART: partial decision trees.

Discussion

Principal Findings

In our previous study, we proposed a simple tool for the prediction of spontaneous resolution and chronic hepatitis C based on IFNL3 genotype and genetic profiles (GUP/GFP) using a combination of HLA-B, HLA-C, and KIR genes. In this study, a data mining methodology was followed to extract relevant hidden features, making it possible to identify subsets of relevant features that would provide greater precision when classifying patients into those who will go on to develop chronic hepatitis C and those who will experience spontaneous resolution of HCV infection. More specifically, in this study we present two models able to predict HCV outcome in each patient using just a subset of features: a simpler, interpretable one using just 4 features and a more complex one using 24 features. The study of both PART and ensemble models demonstrated that they yielded a much better predictive performance than the tool used in the previous study according to a number of different evaluation metrics such as CCR, PPV, NPV, sensitivity, specificity, and AUROC.

The factors analyzed in our previous study using conventional univariate and multivariate analysis showed that there was a strong association between IFNL3, HLA-B*44, KIR3DS1, and HLA-C*12 and the probability of developing chronic hepatitis C [3]. This study confirmed that the same factors were also important for HCV outcome (most of them are included in the PART-1 model and all of them in the ensemble model). In our previous study, however, we did not find any association with KIR2DS1, which is used in both models to predict which class the patient belongs to. This demonstrates that data mining is able to detect complex associations between factors, going beyond the analysis of individual factors commonly used in biomedicine. It is also interesting that the data mining methodology was able to identify genetic patterns hidden in univariate and multivariate analysis on the basis of a total of 138 patients. In the context of GWAS, many studies have found SNPs associated with pathologies without finding the mechanism or molecular basis to explain the associations, since the etiology of most human diseases is multifactorial and involves numerous genes. In this context, the data mining approach could facilitate the discovery of previously hidden genetic patterns in studies with high-dimensional data.

Limitations

There are, however, certain differences in terms of the clinical applicability of the data mining approach depending on the type of model. In this study, for example, using PART, we were able to obtain an interpretable model (as in the previous study) and also a higher predictive performance. With the ensemble model, we obtained a higher predictive performance but lost interpretability (black-box model). Hence, if we are aiming to obtain the best possible predictive performance for patients, we should focus on the ensemble model. If the interpretability of the model is also of interest, since it gives the clinician useful information, the PART model would be preferred. More studies would be necessary to balance the accuracy of models against suitability for implementation in clinical decision making.

Conclusion

Performance of data mining techniques in this study identified genetic patterns that were hidden by the conventional methodology using two models that increased the classification accuracy of HCV outcome. The data mining methodology could be used as an alternative approach in biomedicine, facilitating knowledge in the management of human diseases.

Acknowledgments

This work was supported by the Fundación Progreso y Salud, Consejería de Salud de la Junta de Andalucía (PI0187/2013), the Fundación para la Investigación en Salud del Instituto Carlos III (PI15/01017), the Red de Investigación en SIDA de España ISCIII-RETIC (RD16/0025/0034), the University of Cordoba and Andalusian Council of Economy, Knowledge, Business, and Universities (project 1262678-F), and the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (TIN2017-83445-P). AR is the beneficiary of Contratos para la intensificación de la actividad investigadora en el Sistema Nacional de Salud by the Ministerio de Ciencia, Promoción y Universidades of Spain (INT20-00028). ARJ and MF are recipient of postdoctoral perfection grants by the Ministerio de Ciencia, Innovación y Universidades of Spain (CP18/00111 and CD18/00091, respectively). We would like to thank Mrs Janet Dawson for revising the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Area under the receiver operating characteristic of each model computed as the average value of 30 executions.
[XLSX File (Microsoft Excel File), 24 KB - jmir_v23i2e18766_app1.xlsx]

References

1. Thomas DL, Thio CL, Martin MP, Qi Y, Ge D, O'Huigin C, et al. Genetic variation in IL28B and spontaneous clearance of hepatitis C virus. Nature 2009 Oct 08;461(7265):798-801 [FREE Full text] [doi: 10.1038/nature08463] [Medline: 19759533]
2. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PIW, Chen H, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007 Jun 01;316(5829):1331-1336 [FREE Full text] [doi: 10.1126/science.1142358] [Medline: 17463246]
3. Frias M, Rivero-Juárez A, Rodriguez-Cano D, Camacho ?, López-López P, Risalde M, et al. HLA-B, HLA-C and KIR improve the predictive value of IFNL3 for hepatitis C spontaneous clearance. Sci Rep 2018 Jan 12;8(1):659. [doi: 10.1038/s41598-017-17531-7]
4. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. Int J Med Inform 2008 Feb;77(2):81-97. [doi: 10.1016/j.ijmedinf.2006.11.006] [Medline: 17188928]
5. Abd Elrazek AEMA, Mahfouz HM, Metwally AM, El-Shamy AM. Mortality prediction of nonalcoholic patients presenting with upper gastrointestinal bleeding using data mining. Eur J Gastroenterol Hepatol 2014 Feb;26(2):187-191. [doi: 10.1097/MEG.0b013e328365c3b0] [Medline: 24088733]
6. Grossi E, Buscema M. Introduction to artificial neural networks. Eur J Gastroenterol Hepatol 2007;19(12):1046-1054. [doi: 10.1097/meg.0b013e3282f198a0]
7. Peng Y, Wu Z, Jiang J. A novel feature selection approach for biomedical data classification. J Biomed Inform 2010 Feb;43(1):15-23. [doi: 10.1016/j.jbi.2009.07.008]
8. Chen AH, Lin C. A novel support vector sampling technique to improve classification accuracy and to identify key genes of leukaemia and prostate cancers. Exp Syst App 2011 Apr;38(4):3209-3219. [doi: 10.1016/j.eswa.2010.09.009]
9. Martin-Sanchez F, Verspoor K. Big data in medicine is driving big changes. Yearb Med Inform 2018 Mar 05;23(01):14-20. [doi: 10.15265/iy-2014-0020]
10. Yenduri S, Iyengar SS. Performance evaluation of imputation methods for incomplete datasets. Int J Soft Eng Knowl Eng 2011 Nov 21;17(01):127-152. [doi: 10.1142/s0218194007003173]
11. Classification accuracy of HCV infection outcome: Data mining approach. 2020. URL: http://www.uco.es/kdis/hcv [accessed 2021-01-29]
12. Gahete MD, del Rio-Moreno M, Camargo A, Alcala-Diaz JF, Alors-Perez E, Delgado-Lista J, et al. Changes in splicing machinery components influence, precede, and early predict the development of type 2 diabetes: from the CORDIOPREV Study. EBioMedicine 2018 Nov;37:356-365. [doi: 10.1016/j.ebiom.2018.10.056]
13. Ruiz R, Riquelme J, Aguilar-Ruiz J. Heuristic search over a ranking for feature selection. Proc IWANN 2005 Lect Notes Comp Sci 2005;3512:742-749. [doi: 10.1007/11494669_91]
14. Reyes O, Morell C, Ventura S. Scalable extensions of the ReliefF algorithm for weighting and selecting features on the multi-label learning context. Neurocomputing 2015 Aug;161:168-182. [doi: 10.1016/j.neucom.2015.02.045]

15. Reyes O, Morell C, Ventura S. Evolutionary feature weighting to improve the performance of multi-label lazy algorithms. Integrated Comp Aided Eng 2014 May 29;21(4):339-354. [doi: 10.3233/ica-140468]
16. García S, Luengo J, Herrera F. Intelligent Systems Reference Library: Data Preprocessing in Data Mining. Berlin: Springer; 2015.
17. Salzberg S. Book review: C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. Mach Learn 1994;16:235-240 [FREE Full text] [doi: 10.1007/bf00993309]
18. Witten I, Frank E. Data mining: practical machine learning tools and techniques. Biomed Eng Online 2006;5:51. [doi: 10.1016/C2009-0-19715-5]
19. Dash M, Liu H. Consistency-based search in feature selection. Artif Intell 2003 Dec;151(1-2):155-176. [doi: 10.1016/s0004-3702(03)00079-1]
20. Yang Y, Pedersen J. A comparative study on feature selection in text categorization. ICML '97 Proc 14th Int Conf Mach Learn 1997;97:412-420. [doi: 10.5555/645526.657137]
21. Robnik-Sikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RreliefF. Mach Learn 2003;53:23-69. [doi: 10.1023/a:1025667309714]
22. Hall M. Correlation-based Feature Subset Selection for Machine Learning [Thesis]. Hamilton: University of Waikato; 1998.
23. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. SIGKDD Explor Newsl 2009 Nov 16;11(1):10-18. [doi: 10.1145/1656274.1656278]
24. Frank E, Witten I. Generating accurate rule sets without global optimization. ICML '98 Proc 15th Int Conf Mach Learn 1998:144-151. [doi: 10.5555/645527.657305]
25. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
26. Clemmensen L, Hastie T, Witten D, Ersbøll B. Sparse discriminant analysis. Technometrics 2011 Nov;53(4):406-413. [doi: 10.1198/tech.2011.08118]
27. Hsu C, Chang C, Lin C. A practical guide to support vector classification. Department of Computer Science, National Taiwan University. 2003. URL: http://www.csie.ntu.edu.tw/~cjlin/papers.html [accessed 2021-01-26]
28. Meier L, Van de Geer S, Bühlmann P. The group lasso for logistic regression. J Royal Stat Soc Series B Stat Methodol 2008;70(1):53-71 [FREE Full text] [doi: 10.1111/j.1467-9868.2007.00627.x]
29. Kuhn M. Building predictive models in R using the caret Package. J Stat Soft 2008;28(5):1-26. [doi: 10.18637/jss.v028.i05]
30. Dietterich T. Ensemble methods in machine learning. In: Multiple Classifier Systems. Berlin: Springer; 2000.

Abbreviations

AUROC: area under the receiver operating characteristic
Bw: epitope Bw
CC: genotype CC
CCR: correct classification rate
GFP: genetically favorable profile
GUP: genetically unfavorable profile
GWAS: genome-wide association study
HCV: hepatitis C virus
HLA: human leukocyte antigen
IFNL3: interferon lambda–3
KIR: killer cell immunoglobulin-like receptor
NPV: negative predictive value for chronic hepatitis C
PART: partial decision trees
PPV: positive predictive value for spontaneous resolution
SNP: single nucleotide polymorphism
WEKA: Waikato Environment for Knowledge Analysis

Edited by R Kukafka; submitted 18.03.20; peer-reviewed by E Menasalvas, H Nilsson, P Kocbek; comments to author 11.09.20; revised version received 02.11.20; accepted 17.12.20; published 24.02.21.

Please cite as:
Frias M, Moyano JM, Rivero-Juarez A, Luna JM, Camacho Á, Fardoun HM, Machuca I, Al-Twijri M, Rivero A, Ventura S
Classification Accuracy of Hepatitis C Virus Infection Outcome: Data Mining Approach
J Med Internet Res 2021;23(2):e18766
URL: https://www.jmir.org/2021/2/e18766
doi:10.2196/18766
PMID:33624609

Original Paper

A Clinical Tool (CUE-tool) for Health Care Professionals to Assess the Usability and Quality of the Content of Medical Information Websites: Electronic Delphi Study

Leonie Klompstra[1], PhD[‡]; Maria Liljeroos[1,2], PhD; Johan Lundgren[1], PhD; Brynja Ingadottir[3,4], PhD

[1]Department of Health, Medicine and Caring Sciences, Linköping University, Norrköping, Sweden

[2]Centre for Clinical Research Sörmland, Uppsala University, Eskilstuna, Sweden

[3]Faculty of Nursing, School of Health Sciences, University of Iceland, Reykjavik, Iceland

[4]Landspitali University Hospital, Reykjavik, Iceland

[‡]the CESAR (Collaboration and Exchange in Swedish cardiovascular caring Academic Research) network

Corresponding Author:
Leonie Klompstra, PhD
Department of Health, Medicine and Caring Sciences
Linköping University
Campus Norrköping
Bredgatan 33
Norrköping, 60174
Sweden
Phone: 46 11363629
Email: leonie.klompstra@liu.se

Abstract

Background: As patients are increasingly searching for information about their medical condition on the internet, there is a need for health professionals to be able to guide patients toward reliable and suitable information sources on the internet.

Objective: The aim of the study was to develop a clinical tool for health care professionals to assess the usability and quality of the content of websites containing medical information that could be recommended to patients.

Methods: A 3-round modified electronic Delphi (eDelphi) study was conducted with 20 health care professionals.

Results: In round one of the eDelphi study, of the 68 items initially created, 41 items (29 on usability and 12 on content) were rated as important or very important by more than half of the panel and thus selected for further evaluation in round two. In round two, of the 41 items chosen from round 1, 19 were selected (9 on usability and 10 on content) as important or very important by more than half of the panel for further evaluation. As a result of round three, 2 items were combined as a single item, leaving the instrument with 18 items in total (8 on usability and 10 on content). The tool is freely accessible online.

Conclusions: The CUE-tool can be used to (1) evaluate the usability and reliability of the content of websites before recommending them to patients as a good information source; (2) identify websites that do not have reliable content or may be difficult for patients to use; (3) develop quality websites by using the criteria in the CUE-tool; and (4) identify different qualities between different websites.

(J Med Internet Res 2021;23(2):e22668) doi:10.2196/22668

KEYWORDS

self-care; smartphone; internet; apps; websites; eDelphi

Introduction

Background

Patients are increasingly searching for information about their medical condition on the internet. In high-income countries 75% of the population reported that they search the internet for health information [1,2]. Using the internet as a source for different types of health-related information (eg, reading about medications, reading about other persons' health experiences, watching health-related videos, and signing up for different health email updates) is more common among patients with a chronic illness compared with people without a chronic illness

[2]. Health-related information on the internet is gradually replacing health professionals as important sources of reliable and independent information regarding health and treatment. There are advantages related to online patient education. For example, patients with chronic diseases reported, after seeking information from disease-specific websites, that they were taking their medications more regularly and adhering to treatment to a greater extent [3]. However, when patients lack guidance, there is a risk that their condition becomes worsened by, for example, waiting for too long seeking health care with symptoms of deterioration [4].

Although websites and smartphone apps are available with reliable information, there are also numerous examples of low-quality information regarding health and medical problems accessible through the internet [5]. Furthermore, the information could be very reliable (eg, peer-reviewed open access articles) but less suitable or understandable for the average patient.

Self-care (a rational process involving purposeful choices and behaviors, reflecting knowledge and thought [6]) is essential in the management of most illnesses, and knowledge about own health condition has been identified as an important prerequisite for successful self-care. To gain knowledge, the patient needs access to reliable and understandable information [7]. In order to gain knowledge, whether from the internet or other sources, the patient needs to be able to both read and understand the information [8]. As the information available on the internet is excessive, patients increasingly turn to health care for help with choosing reliable websites [9]. Besides, the information available on websites is not always of acceptable quality or usable for the patient. Therefore, there is a need for health professionals to be able to guide patients toward reliable and suitable information sources on the internet.

An important result of a comprehensive website evaluation [10] was the need for a practical, easy-to-use tool to evaluate websites. There are a number of different tools to evaluate websites; however, these tools seldom consider both the reliability and appropriateness of the information as well as the readability/comprehensibility of the information at the same time.

Purpose

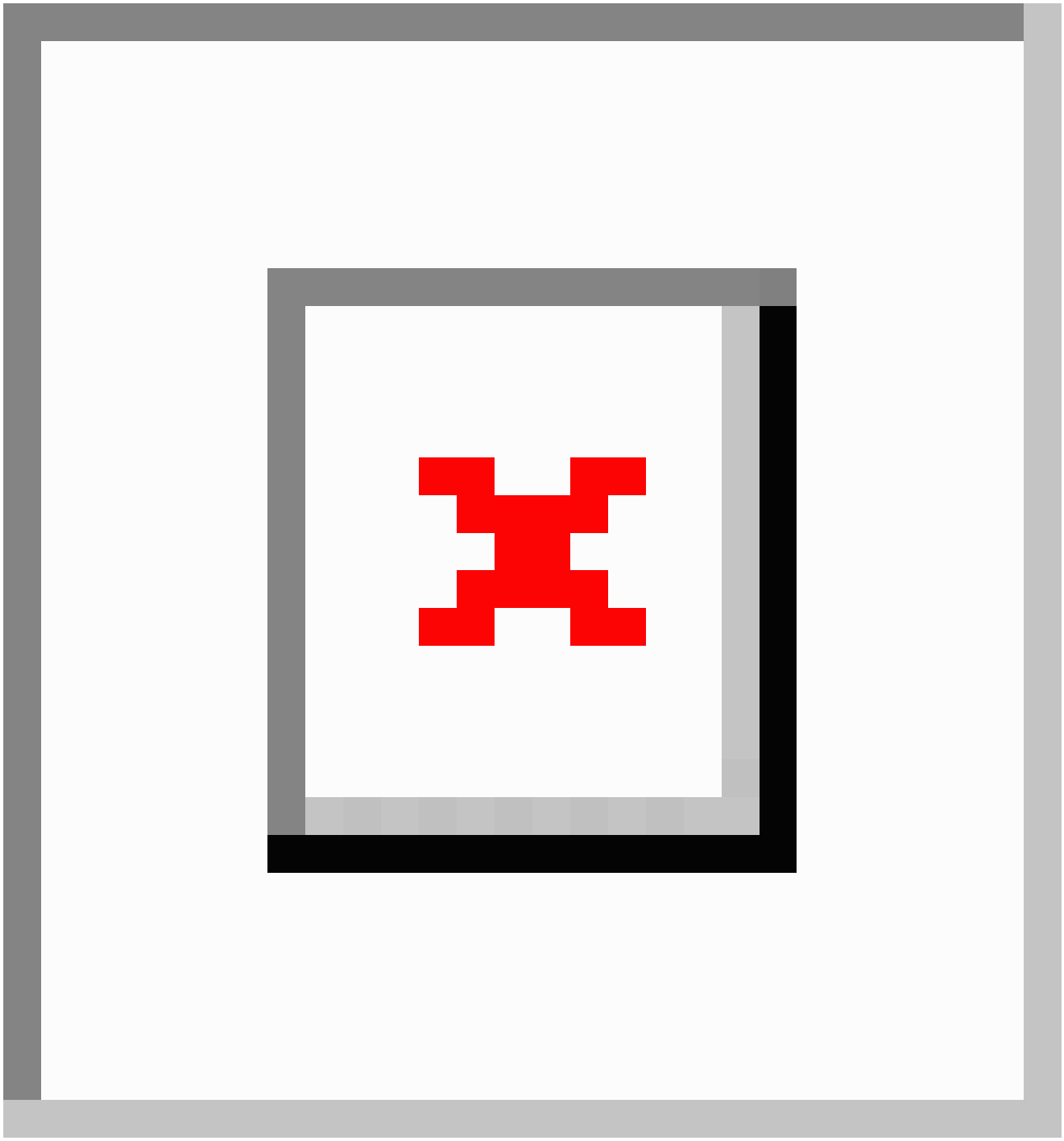
In this study, we aimed to develop a clinical tool for health care professionals to assess the usability and quality of the content of websites containing medical information that could be recommended to patients.

Methods

Electronic Delphi Study

For the development of a tool to assess the quality of websites containing health-related medical information for patients, a 3-round modified electronic Delphi (eDelphi) study was conducted (Figure 1). An eDelphi study is a structured process distributing a series of questionnaires during several rounds to gather information and set priorities or gain consensus regarding a specific issue [11,12]. The eDelphi technique allows the inclusion of a large number of individuals across diverse geographical locations without physically meeting them. The eDelphi technique is often conducted via online web surveys, offering a number of advantages, as they are quick to set up, relatively low cost, and provide high level of data security [13]. Systematic feedback, structured information flow, and iteration and anonymity are the main characteristics of an eDelphi technique. Systematic feedback of experts' responses takes place in-between rounds by informing individual experts about the group opinions. Iteration takes place by presenting feedback via a certain number of rounds [12].

Figure 1. A 3-round eDelphi study for developing the CUE-tool. eDelphi: electronic Delphi.



Procedures and Participants

The panel consisted of health care professionals, selected based on their publications within patient education or that they were members of the CESAR (Collaboration and Exchange in Swedish cardiovascular caring Academic Research) network, a professional research network in Sweden. We aimed for a multidisciplinary panel of health care professionals, with diversity in gender and from variety of countries. The health care professionals that participated in the first round were also approached for the second and third rounds. Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Iceland [14].

Before round one of the study's eDelphi, items on usability and content were created based on available literature on website evaluation and the general information needs of patients. The theoretical perspective of empowering patient education [15] guided the development of the items in the content part of the tool. In empowering patient education, the emphasis is on patients' knowledge expectations and the knowledge patients receive. The more patient expectations of knowledge are met with the received knowledge, the more possibilities there are for empowerment and self-management [15].

A total of 68 items were created for the first round, of which 58 were based on the available literature on tools for evaluation of websites and 10 on content from the multidimensional empowering patient education [15].

The panel was asked to rate the importance of each item on a 4-point scale (response options from 1 [Very important] to 4 [Not important at all]) and invited to comment on each item.

Items were selected for round two if more than half of the participants rated them as important or very important and the same method was applied for round three. In round three the panel was asked to answer 5 additional questions on the developed tool:

1. Are there any items missing?

2. Are there any items unclear?

3. Please look at the examples given in the tool. Do you agree on the examples or do you have suggestions for better examples?

4. Please give your thoughts about the end product of the tool.

5. Do you have a suggestion on a name of the instrument that is easy to use and reflects the area of use?

After the development of the tool, it was translated from English to Swedish, Icelandic, and Dutch and its face feasibility tested by nurses and allied professionals at the EuroHeartCare Conference 2019, Icelandic Nurses' Association Conference Hjúkrun 2019, Icelandic nursing students, and in 2 primary care centers and 1 in-hospital heart failure clinic in Sweden. These health care professionals could choose any website including patient information, in any language, and they were asked if the aim of the development of the CUE-tool was met and their opinion about the usability of the scale. All of the health care professionals agreed that the aim was met, and they were positive about using the tool to evaluate websites. No adjustments were needed.

Results

In total, 20 out of 34 health care professionals, from 5 countries, responded to the invitation to participate in the study in round one and 18 of them responded in rounds two and three. The panel came from Sweden, the Netherlands, Iceland, Norway, and Finland. The panel members' age ranged from 30 to 69 (mean 46 [SD 11]) years and 15/20 were female (75%). Among the panel members, 13 had a doctorate, 4 had a master's degree, and 1 had a bachelor's degree (Table 1).

Table 1. Electronic Delphi (eDelphi) panel demographics.

Demographics	Round 1 (N=20)
Female, n (%)	15 (75)
Age (years), mean (SD)	46 (11)
Education, n (%)	
Doctorate	13 (65)
Master's degree	4 (20)
Bachelor's degree	1 (5)
Missing	2 (10)
Experience in patient education (years), mean (SD)	15 (10)
Main area of work role, n (%)	
Clinical	1 (5)
Research	12 (60)
Clinical and research	6 (30)
Other	1 (5)

In round one of the eDelphi study, of the 68 items initially created, 41 (29 on usability and 12 on content) were rated as important or very important by more than half of the panel and selected for further evaluation in round two (Figure 1).

In round two, of the 41 items chosen from round 1, 19 were selected (9 on usability and 10 on content) as important or very important by more than half of the panel, and thus these items were considered for further evaluation.

As a result of round three, 2 items were put together as a single item, leaving the instrument with 18 items in total (Multimedia Appendix 1). The panel did not miss any items in the tool, and all the items were clear. Examples given in the instrument to clarify items were also agreed on.

As an end product, the panel preferred a scoring system with 2 separate summative scores, 1 for usability and 1 for content, ranging from 0 to 100. From the suggestions made by the panel on a name for the tool, the authors chose the *CUE-tool* as an acronym for "The Credible and Usable Evaluation of patient education tool for web-sites."

The nurses and allied professionals (N = 100) who tested the CUE-tool (on paper) when evaluating websites for patient educational purposes had no problems using it. All items were clear and the tool was seen as an addition to practice. To have the tool online with a summative scoring system was seen as an asset. The tool is freely accessible online [16] and a copy of the tool is also presented as Multimedia Appendix 1. A summative scoring and reliability assessment will be performed in the future.

Discussion

Because patients are increasingly searching for information about their medical condition online, it was important to develop a tool for health care professionals so they can advise patients on suitable and reliable websites from which they can seek disease-related information.

The CUE-tool is an easy-to-use website evaluation tool that helps the user to evaluate both the website's usability and quality of its content. A recent review [17] evaluated patients' preferences for the design features of an effective online education website and found that the information should be patient tailored, interactive, and the content credible and clearly presented. Patients also found multimedia and high interpretability to be essential design features of online patient education websites for chronic disease management. All these features are assessed when using the CUE-tool.

The composition of the eDelphi panel may have affected the results of this study and the development of the tools because the majority were researchers, although of different backgrounds (age and profession). However, most of them had previous work experience as clinical nurses at hospitals, in both wards and outpatient clinics. Although we used patient knowledge expectations while developing the content items in the CUE-tool by using the theoretical perspective of empowering patient education [15], we did not include patients in the eDelphi panel. This decision was made because the tool was designed to be used as a clinical tool to assist health care professionals in finding websites that could be recommended to patients. Although the CUE-tool is developed to be used as a clinical tool, it is also useful outside of health care, for example, in patient organizations or in developing websites for patient education.

The use of empowering patient education [15] as the theoretical foundation of the content items and performing an eDelphi study to develop the CUE-tool may provide specific future possibilities of its applications in (nursing) research and education besides its clinical usefulness. For example, the CUE-tool can be used to (1) evaluate the usability and reliability of the content of websites before recommending them to patients as a good information source; (2) identify websites that do not have reliable content or may be difficult for patients to use; (3) develop quality websites using the criteria in the CUE-tool; (4) identify different qualities between different websites. Accurate and detailed assessments of available websites providing health information can be a valuable resource in teaching strategies to increase knowledge and self-care of patients. Accordingly, health care professionals can create new teaching interventions and revise curricula based on reliable websites identified in these assessments. For further validation, we will include patients' perspectives and the reliability of the scoring system of the CUE-tool will be assessed in future research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The CUE-tool: The Credible and Usable Evaluation of patient education tool for websites.
[PDF File (Adobe PDF File), 136 KB - jmir_v23i2e22668_app1.pdf]

References

1. Bujnowska-Fedak M, Waligóra J, Mastalerz-Migas A. The Internet as a Source of Health Information and Services. Adv Exp Med Biol 2019;1211:1-16. [doi: 10.1007/5584_2019_396] [Medline: 31273574]
2. Madrigal L, Escoffery C. Electronic Health Behaviors Among US Adults With Chronic Disease: Cross-Sectional Survey. J Med Internet Res 2019 Mar 05;21(3):e11240 [FREE Full text] [doi: 10.2196/11240] [Medline: 30835242]
3. Dolor RJ, Yancy WS, Owen WF, Matchar DB, Samsa GP, Pollak KI, et al. Hypertension Improvement Project (HIP): study protocol and implementation challenges. Trials 2009 Mar 26;10(1):13 [FREE Full text] [doi: 10.1186/1745-6215-10-13] [Medline: 19245692]
4. Vasaroangrong T, Thosingha O, Riegel B, Ruangsetakit C, Viwatwongkasem C. Factors influencing prehospital delay time among patients with peripheral arterial occlusive disease. Eur J Cardiovasc Nurs 2016 Jun 20;15(4):285-293. [doi: 10.1177/1474515114567813] [Medline: 25604723]
5. Keselman A, Arnott Smith C, Murcko AC, Kaufman DR. Evaluating the Quality of Health Information in a Changing Digital Ecosystem. J Med Internet Res 2019 Feb 08;21(2):e11129 [FREE Full text] [doi: 10.2196/11129] [Medline: 30735144]
6. Riegel B, Carlson B, Moser DK, Sebern M, Hicks FD, Roland V. Psychometric testing of the self-care of heart failure index. J Card Fail 2004 Aug;10(4):350-360. [doi: 10.1016/j.cardfail.2003.12.001] [Medline: 15309704]
7. Riegel B, Jaarsma T, Strömberg A. A middle-range theory of self-care of chronic illness. ANS Adv Nurs Sci 2012;35(3):194-204. [doi: 10.1097/ANS.0b013e318261b1ba] [Medline: 22739426]
8. Speros C. Health literacy: concept analysis. J Adv Nurs 2005 Jul;50(6):633-640. [doi: 10.1111/j.1365-2648.2005.03448.x] [Medline: 15926968]

9. Ingadottir B, Blondal K, Jaarsma T, Thylen I. Perceptions about traditional and novel methods to learn about postoperative pain management: a qualitative study. J Adv Nurs 2016 Dec 13;72(11):2672-2683. [doi: 10.1111/jan.13021] [Medline: 27205900]
10. Klompstra L, Johansson Östbring M, Jaarsma T, Ågren S, Fridlund B, Hjelm C, et al. The Appropriateness and Presentation of Commonly Available Cardiovascular Web Pages Providing Information About Cardiovascular Diseases. CIN: Computers, Informatics, Nursing 2019;37(10):493-497. [doi: 10.1097/cin.0000000000000595]
11. Keeney S, Hasson F, McKenna H. Consulting the oracle: ten lessons from using the Delphi technique in nursing research. J Adv Nurs 2006 Jan;53(2):205-212. [doi: 10.1111/j.1365-2648.2006.03716.x] [Medline: 16422719]
12. Keeney S, Hasson F, McKenna H. The Delphi Technique in Nursing and Health Research. Chichester, West Sussex, UK: John Wiley & Sons; 2011.
13. Gill FJ, Leslie GD, Grech C, Latour JM. Using a web-based survey tool to undertake a Delphi study: application for nurse education research. Nurse Educ Today 2013 Dec;33(11):1322-1328. [doi: 10.1016/j.nedt.2013.02.016] [Medline: 23510701]
14. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 May;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
15. Valkeapää K, Klemetti S, Cabrera E, Cano S, Charalambous A, Copanitsanou P, et al. Knowledge expectations of surgical orthopaedic patients: a European survey. Int J Nurs Pract 2014 Dec;20(6):597-607. [doi: 10.1111/ijn.12189] [Medline: 24118436]
16. CUE-Tool. URL: https://liu.se/en/research/cesar-network/cue-verktyget [accessed 2021-02-07]
17. Win KT, Hassan NM, Oinas-Kukkonen H, Probst Y. Online Patient Education for Chronic Disease Management: Consumer Perspectives. J Med Syst 2016 May 4;40(4):88. [doi: 10.1007/s10916-016-0438-0] [Medline: 26846749]

Abbreviations

CESAR: Collaboration and Exchange in Swedish cardiovascular caring Academic Research
CUE-tool: The Credible and Usable Evaluation of patient education tool for websites

Edited by G Eysenbach; submitted 20.07.20; peer-reviewed by N Sharifzadeh, J Taylor; comments to author 17.10.20; revised version received 30.11.20; accepted 24.12.20; published 17.02.21.

Please cite as:
Klompstra L, Liljeroos M, Lundgren J, Ingadottir B
A Clinical Tool (CUE-tool) for Health Care Professionals to Assess the Usability and Quality of the Content of Medical Information Websites: Electronic Delphi Study
J Med Internet Res 2021;23(2):e22668
URL: http://www.jmir.org/2021/2/e22668/
doi:10.2196/22668
PMID:33595439

Original Paper

Opening a “Can of Worms” to Explore the Public's Hopes and Fears About Health Care Data Sharing: Qualitative Study

Olivia Lounsbury[1,2*], BS; Lily Roberts[1*], BSc; Jonathan R Goodman[1], MA, MPhil; Philippa Batey[3], MRes; Lenny Naar[3], BFA; Kelsey M Flott[1], BA, MSc, PhD; Anna Lawrence-Jones[1], BSc (Hons); Saira Ghafur[4], MBChB, MRCP (Resp), MSc; Ara Darzi[1], PC, KBE, FRS, FMedSci, HonFREng; Ana Luisa Neves[1,5], MSc, PhD, MD

[1]Patient Safety Translational Research Centre, Institute of Global Health Innovation, London, United Kingdom

[2]Patient Safety Movement Foundation, Irvine, CA, United States

[3]The Helix Centre, Institute of Global Health Innovation, London, United Kingdom

[4]Centre for Health Policy, Institute of Global Health Innovation, Imperial College London, London, United Kingdom

[5]Center for Health Technology and Services Research / Department of Community Medicine, Health Information and Decision, Faculty of Medicine, University of Porto, Porto, Portugal

[*]these authors contributed equally

Corresponding Author:
Ana Luisa Neves, MSc, PhD, MD
Patient Safety Translational Research Centre
Institute of Global Health Innovation
Imperial College London, St Mary’s Campus
Queen Elizabeth Queen Mother Wing
London, W2 1NY
United Kingdom
Phone: 44 (0)20 7589 5111
Email: ana.luisa.neves14@imperial.ac.uk

Abstract

Background: Evidence suggests that health care data sharing may strengthen care coordination, improve quality and safety, and reduce costs. However, to achieve efficient and meaningful adoption of health care data-sharing initiatives, it is necessary to engage all stakeholders, from health care professionals to patients. Although previous work has assessed health care professionals’ perceptions of data sharing, perspectives of the general public and particularly of seldom heard groups have yet to be fully assessed.

Objective: This study aims to explore the views of the public, particularly their hopes and concerns, around health care data sharing.

Methods: An original, immersive public engagement interactive experience was developed—*The Can of Worms* installation—in which participants were prompted to reflect about data sharing through listening to individual stories around health care data sharing. A multidisciplinary team with expertise in research, public involvement, and human-centered design developed this concept. The installation took place in three separate events between November 2018 and November 2019. A combination of convenience and snowball sampling was used in this study. Participants were asked to fill self-administered feedback cards and to describe their hopes and fears about the meaningful use of data in health care. The transcripts were compiled verbatim and systematically reviewed by four independent reviewers using the thematic analysis method to identify emerging themes.

Results: Our approach exemplifies the potential of using interdisciplinary expertise in research, public involvement, and human-centered design to tell stories, collect perspectives, and spark conversations around complex topics in participatory digital medicine. A total of 352 qualitative feedback cards were collected, each reflecting participants’ *hopes* and *fears* for health care data sharing. Thematic analyses identified six themes under *hopes*: enablement of personal access and ownership, increased interoperability and collaboration, generation of evidence for better and safer care, improved timeliness and efficiency, delivery of more personalized care, and equality. The five main *fears* identified included inadequate security and exploitation, data inaccuracy, distrust, discrimination and inequality, and less patient-centered care.

Conclusions: This study sheds new light on the main hopes and fears of the public regarding health care data sharing. Importantly, our results highlight novel concerns from the public, particularly in terms of the impact on health disparities, both at international

and local levels, and on delivering patient-centered care. Incorporating the knowledge generated and focusing on co-designing solutions to tackle these concerns is critical to engage the public as active contributors and to fully leverage the potential of health care data use.

(J Med Internet Res 2021;23(2):e22744) doi:10.2196/22744

KEYWORDS

electronic health records; patient participation; data sharing; patient safety; data security

Introduction

Background

With the advent of the digital age, health care professionals have witnessed significant advancements in innovation and research. The increasing adoption of digital technologies and electronic health records (EHRs) has expanded the capacity for interoperability and data sharing, both for primary uses (ie, direct care) and secondary uses (ie, research, public health, health policy). Evidence suggests that health care data sharing may strengthen care coordination, improve quality and safety, and reduce costs [1]. However, the growing complexity of systems, stakeholders, and unbounded ecosystems, unsuccessful data sharing initiatives (eg, the care.data program in the United Kingdom), and recent cybersecurity incidents (eg, the WannaCry attack) and evolving regulations (eg, General Data Protection Regulation) [2] have contributed to the increasing lack of clarity and trust by the public [3]. Therefore, data sharing is becoming an increasingly controversial subject, with many researchers and patients reporting concerns about how and why health care data are shared [4]. Previous evidence highlights that the most common concern is patient privacy [5,6]; even when data are anonymized, there remains a risk that by using only a few data points, patients can be reidentified by their own health information [7]. Previous studies also highlight that public support is generally higher when data are used for the *greater good*, but the acceptance rates fall steeply when data are shared for use by commercial companies [8].

In recent years, several health care data sharing initiatives have been implemented globally. HealthData Research UK, as part of the Industrial Strategy Challenge Fund, has launched 7 new data hubs as part of a 4-year program to create a UK-wide system for the secure and safe use of large-scale health data [9]. In the United States, the Institute for Healthcare Improvement's Triple Aim Initiative uses geographic health data to comprehensively understand population health by location, with the aim of improving the experience of care, health of populations, and cost-effectiveness [10]. Innovative engagement programs, as opposed to traditional banners, posters, and advertisements, are especially promising for promoting an informed choice [11]. Over the past several years, social media campaigns and even live theatrical performances have been used to improve the understanding of data sharing practices [12,13]. However, to achieve efficient and meaningful adoption of health care data sharing initiatives, it is necessary to engage all stakeholders, including policy makers, health care professionals, researchers, patients, and the public. A broad understanding of their views, hopes, and concerns about data sharing is crucial to frame the breadth of perspectives, increase adoption, support progress, and enhance equity of care delivery.

Previous research has largely focused on health care provider perspectives and found that providers hope data sharing can have a positive impact by tailoring and improving care delivery [14-17]. Previous research addressing health care professionals' views identified several benefits of data sharing: improved population health, ease of access, and reduced costs [17-19]. However, talking about data sharing is like opening a metaphorical can of worms: when the subject is brought up, many concerns also emerge, including issues around patients' willingness to share, trust, privacy, transparency, confidentiality, and security [20].

If data are anonymized, many patients are comfortable with sharing for the improvement of health services and care [21], with as many as 88% of patients trusting the United Kingdom National Health Service (NHS) to store data safely and use it for ethical, research-oriented reasons [22]. Conversely, a recent survey showed that only one-tenth of people would share data willingly with a tech company [23]. Transparency seems to be a critical factor for patients' willingness to share their data, as the more transparent the organizations are with the public about the use of health care data, including, but not limited to, who has access, the rights to access, and the safeguarding processes in place, the greater the public support for data sharing initiatives [24]. However, a deeper exploration of the factors that contribute to the willingness to harness health care data from the general public perspective is still lacking.

In addition, policy and structural changes are necessary to promote a culture of safety and transparency in organizations across the continuum of care [25]. These changes must be standardized and embedded at every level of care, including primary care, secondary care, and community services at local and national levels, to ensure integrity and alignment, and should be guided by international quality standards. Understanding and integrating the public's hopes and concerns into these policy and structural changes is fundamental to ensure the development of patient-centered data sharing policies.

Objectives

This qualitative study aims to explore public views, particularly their hopes and concerns, around health care data sharing.

Methods

Overview

We developed an original, immersive public engagement interactive experience—*The Can of Worms* installation—with which we aimed to challenge the conventions of how members

of the public receive and give information, using interdisciplinary expertise in research, public involvement and human-centered design, to tell stories, collect perspectives, and spark conversation around complex topics.

To meet the aims and objectives of this study, a qualitative descriptive approach was adopted. Qualitative methods were chosen because of their ability to capture descriptive data on individual perceptions, attitudes, and behaviors [26]. A multidisciplinary team including medical doctors (AN, SG, and AD), health service researchers (JG, LR, KF, and OL), designers (PB and LN), and a patient and public involvement specialist (AJ) with previous experience in qualitative research designed the *Can of Worms* concept and performed this study. Members of the public were involved in recruitment, developing the *Can of Worms* concept and reviewing materials.

Recruitment

A combination of convenience and snowballing sampling was used, and no other exclusion criteria (apart from age) were adopted, to optimize the diversity of the sample. Members of the public (<18 years) were invited to participate through a combination of recruitment approaches, including public advertising through posters, distribution of flyers, partners' networks, social media, and word of mouth. The research team had no established relationships with the participants before the event. Free tea and coffee were offered to the participants; however, no financial incentives were provided for participation.

Concept Development

The *Can of Worms* metaphor was used as a novel way of bringing interest to the subject matter. The exhibition design employed a multidisciplinary team of designers, public members, and researchers to craft a user-centered experience for participants. The team prioritized flexibility, inclusivity, and engagement, principles that guided the design process.

People were invited to explore a free (no admission fee) and immersive installation (ie, in which the space was designed to impact the experience for visitors, as described below). Participants were prompted to reflect on the subject of health care data sharing through storytelling and conversation. Visitors were given a *can opener* and an information leaflet and were encouraged to open cans and listen to stories about health care data sharing, stored in a recorder within the can (Figure 1). Each of the 28 fictional stories fit under 1 of 7 categories: diagnosis, individual care, planning, policy, social services, treatment and prevention, and understanding disease. Topics included international data sharing among patients with chronic diseases, genomic analysis, and data-powered predictive algorithms. The fictional stories were developed by researchers based on anecdotal evidence of risks and benefits of data sharing to represent a balanced view. They were further reviewed by clinicians and lay partners to ensure that they were relevant, factual, and realistic, not harrowing, and in plain English. Stories were recorded by a variety of actors and public members to ensure that they were engaging and relatable. Large text-printed versions were also available for anyone with specific access needs. Examples of stories can be found in Multimedia Appendix 1 [27] and a full list of all stories used within the *Can of Worms* installations can be found at the project website [27].

Figure 1. Materials used in the *Can of Worms* events. (a) Cans containing audio stories focusing on a particular aspect of healthcare data sharing. (b) Blank response card used to collect participants' hopes and concerns towards data sharing. (c) Overview of the *Can of Worms* public engagement immersive installation.

Data Collection

Once participants finished listening to some of the recorded stories, they were asked to provide the following information on a card: age bands, hopes, and fears about the meaningful use of data in health care (Figure 1).

The first exhibition took place at the Helix Pop-up within St. Mary's Hospital in Paddington, London, in November 2018, which meant that the members of the public entering the space were mostly patients, family members, medical students, and passersby near the train station. To encourage the participation of people from seldom heard groups (eg, those from deprived backgrounds and minority ethnic groups), a free bus service was organized from White City and Woodlane to the event. The second was held over a weekend at The Great Exhibition Road Festival, South Kensington, in June 2019—*Can of Worms* was one of many exhibits at an event attracting 20,000 members of the general public, as well as students and staff. The third took place on one day at the NHS Digital Academy Residential in Tower Hill in November 2019—this event was open to participants on an academic program for digital health leaders.

Participants' responses were compiled verbatim and were not returned to the participants for comments and/or corrections. As anonymous, self-administered cards were used, only the participant was present when registering data. Participants had minimal knowledge of the research team; thus, the potential for bias and assumptions was kept to a minimum. No repeat interviews were conducted.

Data Analysis

The transcripts were compiled verbatim and systematically reviewed by four independent reviewers using the thematic analysis method to identify emerging themes. The themes were supported by participants' quotations. The Consolidated Criteria for Reporting Qualitative Research were used to ensure that the study met the recommended standards for qualitative data reporting [28].

Ethics

The project was reviewed by the Health Research Authority (HRA) Public Involvement Team HRA. Additional HRA approval by the NHS Research Ethics Committee review was not deemed necessary [29].

Results

Participant Characteristics

A total of 352 participants filled the response card (iteration 1 [Helix Centre]: n=175; iteration 2 [Exhibition Road Festival]: n=142; iteration 3 [NHS Digital Academy]: n=35). In iteration 1, the most frequent age band was 25 to 34 years (51/175, 29.1%). In iteration 2, the most frequent age band was below 18 years (55/142, 38.7%), and in iteration 3, the most frequent age band was 45 to 54 years (27/35, 77%).

The themes presented are listed in no particular order, and in line with the qualitative approach, no one theme is presented as more important than the other.

Hopes

The level of content and detail varied greatly between cards. Thematic analysis of the patients' narratives revealed six emerging overall hopes: (1) enablement of personal access and ownership, (2) increased interoperability and collaboration, (3) generation of evidence for better and safer care, (4) improved timeliness and efficiency, (5) delivery of more personalized care, and (6) equality (Textbox 1).

Textbox 1. Hopes regarding health care data sharing.

Theme 1: personal access and ownership

- "Patients will hold their data and will share what they want with who they want" (ID 155)
- "Data will be there for every patient when they need it and it will follow them" (ID 144)
- "Having your medical data available on your mobile/smart watch could save many lives in an emergency in the future" (ID 65)

Theme 2: interoperability and collaboration

- "That we treat all data collected across the NHS as one, rich resource. At the moment, there are hundreds of data controls in the NHS and sharing is difficult" (ID 160)
- "For seamless, secure sharing of data between patients, GPs, A&E… so that patients can get the best care possible" (ID 37)
- "As a doctor, I find it difficult to provide the best care for my patients without full access to their past medical history, previous imaging and tests, and up to date medication lists. Data sharing between trusts and GPs" (ID 20)
- "The data can be shared to ensure that there is access to information, especially out of normal hours, so that clinicians can always make informed decisions" (ID 149)
- "That all my healthcare data is available to any medical professional I see! To save me time and inaccuracy" (ID 184)
- "Data sharing means I don't have to tell my story again and again" (ID 185)
- "A more open and collaborative approach to healthcare across the globe" (ID 183)

Theme 3: evidence for better and safer care

- "Understandings that will help the whole healthcare sector from big data analysis" (ID 13)
- "Analysing huge amounts of data could help to find out what causes different diseases (...)" (ID 33)
- "I hope keeping and sharing patients' data will lead to more efficient and accurate diagnoses avoiding human error and be able to draw on a bigger database that slips the human mind" (ID 186)
- "Data can be enhanced to improve outcomes for patients [via] better and quicker diagnosis" (ID 187)
- "Provides an evidence base for identifying effective treatments" (ID 143)
- "Analysing huge amounts of data could help to (...) see and predict how epidemic illnesses are spreading" (ID 33)
- "Less likely for mistakes to happen" (ID 55)
- "Investment and development of technology to achieve parity with industry with human factors and evidence-based design and implementation" (ID 152)

Theme 4: timeliness and efficiency

- "That data gets where it needs to at the time it needs to. Patient care is supported and improved by timely access to the right information" (ID 165)
- "(...) an ecosystem that provides pace and a streamlined service – think how much quicker we could help people" (ID 182)
- "Diseases can be detected early on" (ID 73)
- "Broad sharing and easy access of data to help in a quicker understanding of a healthcare issue and [potential cure]" (ID 4)
- "It creates a clear picture of who I am – so I can be helped better and minimize waste in the health service" (ID 180)

Theme 5: personalized care

- "That it will bring a smarter, more cohesive, personalized care" (ID 179)
- "Reduction in anxiety to have to tell your story [repeatedly]" (ID 157)
- "Improving preventive behavior through personalized interventions" (ID 148)

Theme 6: equality

- "That greater ownership by patients of their data will encourage conversations of equality between them and their healthcare providers" (ID 181)
- "The patient data can be used and shared more effectively, for example, [in the care of] transgender patients, the proper pronouns can be used" (ID 9)
- "Data is a very powerful way to tackle inequalities and improve the level of care" (ID 87)

Personal Access and Ownership

Participants hoped that health care data sharing will enable patients access to their own medical records, improving their sense of ownership and involvement in their health and care. Participants also highlighted that improved accessibility of health care data by the patient could prove pivotal in improving safety in medical emergency situations.

Interoperability and Collaboration

The opportunity for enhanced interdisciplinary engagement across the health care field was another hope identified in this study. Participants from each iteration hoped for a more united health care system, ease of collaborative care, and fortified capacity for health care providers to make informed decisions anywhere at any time. Participants hoped for the treatment of data collected as a rich resource instead of the current, fragmented state, which compromises quality and safety of care.

Evidence for Better and Safer Care

Participants also highlighted that data sharing in health care can contribute to providing better and safer care. A significant number of participants from all iterations hoped that data sharing could pave the way for analytical and data mining approaches to improve clinical knowledge in several aspects, including understanding etiology and improving diagnosis and effectiveness of treatment. Some participants also acknowledged that health care data sharing can help understand and monitor the epidemiological nature of certain diseases.

Timeliness and Efficiency

Participants expressed how they hoped that health care might become more efficient as a result of data sharing, as data may be accessed anytime, anywhere. Some saw this from the angle that waste could be minimized and more patients helped if data were shared more widely.

Personalized Care

Another theme emerging from the responses was the hope that health care data sharing would lead to more personalized care. Participants highlighted that it could reduce the anxiety produced by having to tell their story repeatedly due to a lack of integrated health records. Data sharing could also result in a more cohesive health record that could allow care to be tailored to individual needs.

Equality

Participants expressed their hope that data sharing would allow patients to be treated equally, regardless of their backgrounds, predispositions, access to public care, or financial means for private care.

Fears

Regarding the main fears, the thematic analysis of the public's narratives revealed five emerging themes: (1) inadequate security and exploitation, (2) data inaccuracy, (3) distrust, (4) discrimination and inequality, and (5) less patient-centered care (Textbox 2).

Textbox 2. Fears regarding health care data sharing.

Theme 1: inadequate security and exploitation

- "Issues around privacy of data" (ID 194)
- "This data being used by companies to discriminate and make [profit]" (ID 37)
- "That conclusions are made without proper examination of the data [and] that the data is used for nefarious purposes" (ID 143)
- "That private companies could use this data purely for targeting the public health service, thereby driving up the costs" (ID 163)

Theme 2: data inaccuracy

- "Incorrect data" (ID 161)
- "Errors that could corrupt the data" (ID 158)

Theme 3: distrust

- "Horror stories delay inevitable progress" (ID 153)
- "That just one rotten egg will set us back years and we miss out on all the progress that could be made" (ID 196)

Theme 4: discrimination and inequity

- "Could widen inequalities for countries that can't afford these technologies" (ID 189)
- "Introduces bias and fails to support a patient-centered approach in a mental health and community setting" (ID 147)
- "Bias gets perpetuated" (ID 150)
- "That if my serious and ongoing medical conditions get out that it would limit or otherwise negatively impact my career pathway and job options" (ID 178)

Theme 5: less patient-centered care

- "The 'person' is being lost and replaced with numbers" (ID 195)
- "We become too reliant on data and forget about the individual patient" (ID 169)
- "We no longer have conversations with healthcare professionals... many may be replaced by exchange of statistics" (ID 30)
- "Artificial intelligence is seen as a quick and cheap fix and patients get substandard care" (ID 159)

Inadequate Security and Exploitation

Participants conveyed their concerns that health care data sharing could be associated with a lack of privacy and security and would therefore be potentially used for nefarious purposes. Specifically, individuals feared the potential for private companies (ie, pharmaceutical companies) to leverage the data for profit at the expense of the public.

Data Inaccuracy

Data accuracy was also a concern, with some participants expressing worries about incorrect data in their records or computing mistakes that may corrupt their data. Participants expressed worry about the accuracy of communication between clinicians and were concerned that the overreliance on data might further compromise communication.

Distrust

Participants expressed a wide variety of perspectives that shared an overall feeling of distrust and apprehension about the potential for sustained and adequate adoption. Some were wary of the impact of previous negative experiences and how they may impact future initiatives of health care data sharing.

Discrimination and Inequity

Participants expressed concerns that increased health care data sharing would only be possible in better-connected regions, and this could widen the gap between these regions and those that do not have the resources to implement such technologies. On an individual level, participants writing from a patient's perspective were concerned that, if shared widely, certain details of their personal data may introduce or perpetuate biases.

Less Patient-Centered Care

Some participants were concerned about a negative impact on the delivery of care that is respectful of, and responsive to, individual patient needs, preferences, and values. Fears were expressed around health care, becoming too focused on data, with a negative impact on the patient-doctor relationship, communication, and quality of care received. Some were concerned that a strong focus on data and artificial intelligence could have a negative impact on patient centeredness.

Discussion

Principal Findings

The six main *hopes* that participants had for health care data sharing concerned (1) enablement of personal access and ownership, (2) increased interoperability and collaboration, (3)

generation of evidence for better and safer care, (4) improved timeliness and efficiency, (5) delivery of more personalized care, and (6) equality. The five main *fears* that participants expressed in relation to health care data sharing were (1) inadequate security and exploitation, (2) data inaccuracy, (3) distrust, (4) discrimination and inequality, and (5) less patient-centered care.

Findings as Compared With Previous Studies

In this study, participants highlighted personal access and ownership over their health records as a key *hope* for data sharing. This sentiment has been expressed in previous studies; a public engagement exercise by the Wellcome Trust (2010) found that 92% of adults and 97% of young people surveyed supported patient access to their own health records [30]. Furthermore, a recent meta-analysis has shown that providing patients with access to their own records can improve several aspects of quality of care, particularly improving health outcomes and patient safety [31]. The growing body of evidence supporting policies that support data sharing with patients also raises important questions about equity and whether these interventions may exacerbate health inequities by improving outcomes for patients with access to their health care data, while further excluding those with low health literacy or poor access to digital technologies [32,33].

Interoperability and collaboration also emerged as key hopes for data sharing in this study and were described from various perspectives as allowing clinicians to make more informed decisions and avoiding the need for patients to repeatedly narrate their clinical information. Interoperability between systems and care settings is recognized as a key aspiration to achieve the full potential of data sharing [34], and it is necessary to engage stakeholders involved (patients, health care professionals, policy makers, and technical companies). Several aspects need to be considered, including the adoption of international standards [35], improved education and awareness of obstacles, and minimizing privacy and cybersecurity issues [34].

The hope that data sharing would provide evidence for better and safer care is in line with the findings of previous studies. A study by O'Brien et al [36] found that 94% of patients surveyed "thought data sharing would help their doctor to make better decisions about their health". In the last decade, the United Kingdom has witnessed a surge in the secondary use of health care data that has generated population-based evidence to inform the delivery of better care, particularly in the mental health space [37] and prescription patterns [38]. Similarly, in the United States, pilot studies have started predicting readmissions and estimating the risk of complications in newborns [37].

Improved timeliness and efficiency were also identified as key hopes in this study, whether it was for patient benefit (ie, early diagnosis and treatment, improvement in diagnosis) or for improved health care efficiency (ie, allowing care to be delivered faster and to more people). Our findings are in line with previous studies surveying patient perspectives, in which helping make new therapies available faster was one of the 2 most important perceived benefits of data sharing [36].

Importantly, participants hoped that data sharing might contribute to making health care more inclusive and increase transparency around demographic information, particularly gender preference, where the use of proper pronouns is of utmost importance to an individual's identity. Previous research suggests that allowing patients to self-label their gender in their EHRs and specifying their preferred names and pronouns could improve their health care experience [39].

The most common *fear* noted by participants was inadequate security and the exploitation of health data. Following the WannaCry attack of 2017, public confidence in the NHS to handle data has been negatively affected [40]. A web-based patient survey in the United Kingdom found that the most important data sharing risk identified by participants was health data being *stolen by hackers* and that in general they would be more comfortable if they were able to learn how their data were protected [36]. These findings mirror the hopes of health care professionals who acknowledged some of these concerns in previous research [41]. Closely linked with fears regarding inadequate security and exploitation of patient data is the idea that distrust can delay progress and hinder the realization of benefits from data sharing. In line with these findings, previous studies showed that the public opposed data sharing when there were financial gains or profits or the possibility of sharing their health care data with private or commercial companies [42].

Errors that resulted in data inaccuracies were feared by the study participants. Previous studies have also found that the public is concerned that errors in records may be difficult to correct and have a negative impact on their care [40]. Previous research highlights that when using data as evidence for better care, it is important that data quality is prioritized, as only with high quality, clean data to feed artificial intelligence algorithms can meaningful insights be extracted [43].

Fears that increased data sharing would give rise to discrimination and inequity of patients were voiced from different contexts. Our participants were concerned that health disparities would widen between more developed countries and those that could not afford these technologies. Others were concerned that some patients may be unfairly prioritized over others, a point of view that is shared with other studies [40]. In one study, those with a lower socioeconomic status expressed more concerns about data sharing and were less likely to consider the benefits that it offers to society [44]. Discrimination, stigmatization, exploitation, or other repercussions are concerns that have been voiced by participants in similar investigations [42].

Some participants expressed that the increased use of computers and artificial intelligence would diminish patient-centered care. The human touch aspect of being treated by a person rather than a computer was valued by members of the public, and there was concern that this would be attenuated significantly. Although digital solutions can improve patients' safety and efficiency of care, they are not able to replace humanistic skills (ie, compassion, commitment, or empathy) [45].

Strengths, Limitations, and Future Work

This study has several strengths. We used multiple iterations of data collection, coupled with the triangulation of interpretations between researchers with expertise in qualitative research, clinical research, and cognitive science. Data collection was performed in several settings to capture the perspectives, hopes, and concerns of a diverse study sample. This is crucial to inform an equitable approach to increase data sharing in the future.

This study employed a methodologically rigorous approach, leveraging qualitative methods to capture rich, descriptive data on individual perceptions, attitudes, and behaviors [46,47]. In addition, all data collection and analysis were performed according to the Consolidated Criteria for Reporting Qualitative Studies criteria [48].

On a broader scale, the *Can of Worms* installation is replicable and adaptable for different settings and contexts and can be implemented with relative ease for future installments, both for data collection purposes and to enhance awareness and behavioral change in diverse audiences on this subject.

Despite its strengths, this study has some limitations. The sample size (N=352) was small and was sourced from three locations in England. For this reason, and the fact that data sharing sentiments differ between countries, our results may not be representative of the wider UK population or extrapolated to international settings. The attempt to keep the length of the survey short and encourage those who do not normally engage to participate limited the amount of information collected on participant demographics. As contact information was not collected, it was not possible to send the themes back to the participants for feedback. Future research may benefit from involving patients more actively as part of this process, either by allowing them to provide feedback on the findings, or by providing training so that they can actively contribute to the thematic analyses.

However, the minimal request for participant disclosure of information could serve both as a limitation and a strength of this study, as it could also have increased participation rates and reduced information bias. Finally, the quasi-experimental nature of the study, in an attempt to elicit attitudes, may impact the generalizability of the resulting views, as participants were primed on the stories in the cans. The feedback cards were displayed on a feedback wall in the installation. Some people chose to read these before they wrote their feedback, which could have biased their results but, equally, may have prompted them to have a deeper reflection, including other participants' perspectives.

Future work should include methodologically robust quantitative studies focusing on how different factors (demographic and social, patient activation, health, and digital literacy) influence both the general public and professional views on data sharing. Future research should also explore the underlying reasons for the public sentiments expressed by collecting additional insights from a range of study participants. Therefore, this study can serve as a first step to unveil areas for future research, from which more actionable conclusions can be drawn. In addition, future work might benefit from international projects assessing data sharing perspectives, as this may help anticipate possible challenges and solutions before future translational implementation of data sharing mechanisms. Finally, future research may consider assessing social media responses to the installations in addition to qualitative responses based on prompts to highlight similarities and contrasting perspectives based on the feedback mechanism.

Conclusions

In the broader context of sharing health care data, involving the public is critical to create a patient-centric culture in health care systems [49]. This study sheds new light on the main hopes and fears of a sample of the UK public regarding health care data sharing. Importantly, our results highlight novel concerns from the public, particularly regarding the impact of health disparities on delivering patient-centered care. Incorporating the knowledge generated and focusing on co-designing solutions to tackle these concerns are critical to engage the public as active contributors to this decision-making process and to fully leverage the potential of health care data use.

Acknowledgments

Public engagement exhibitions were supported by Google Health (previously known as DeepMind Health), the National Institute for Health Research (NIHR), the NHS England, Understanding Patient Data, Reform, the British Heart Foundation, NHS Digital, and the Imperial College Healthcare NHS Trust. AN, KF, AJ, LR, and OL are funded by the Imperial NIHR Patient Safety Translational Research Centre, with infrastructure support from the Imperial NIHR Biomedical Research Centre. OL was also funded by the Patient Safety Movement Foundation. The authors would like to thank the patient and the public members involved in the development and delivery of the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Two examples of stories used in the Can of Worms events. The stories were read by amateur actors, volunteer patient advocates, and researchers. A more detailed overview of the stories is provided in the project website.
[PDF File (Adobe PDF File), 48 KB - jmir_v23i2e22744_app1.pdf]

Multimedia Appendix 2
COREQ checklist.
[PDF File (Adobe PDF File), 491 KB - jmir_v23i2e22744_app2.pdf]

References

1. Sharing Data, Saving Lives: the Hospital Agenda for Interoperability. American Hospital Association. 2019. URL: https://www.aha.org/system/files/2019-01/Report01_18_19-Sharing-Data-Saving-Lives_FINAL.pdf [accessed 2021-02-10]
2. John B. Are you ready for General Data Protection Regulation? Br Med J 2018 Mar 2;360:k941. [doi: 10.1136/bmj.k941] [Medline: 29500167]
3. Watts G. Data sharing: keeping patients on board. Lancet Digital Health 2019 Nov;1(7):e332-e333. [doi: 10.1016/s2589-7500(19)30163-3]
4. Sterckx S, Rakic V, Cockbain J, Borry P. 'You hoped we would sleep walk into accepting the collection of our data': controversies surrounding the UK care.data scheme and their wider relevance for biomedical research. Med Health Care Philos 2016 Jun 18;19(2):177-190 [FREE Full text] [doi: 10.1007/s11019-015-9661-6] [Medline: 26280642]
5. Papoutsi C, Reed JE, Marston C, Lewis R, Majeed A, Bell D. Patient and public views about the security and privacy of Electronic Health Records (EHRs) in the UK: results from a mixed methods study. BMC Med Inform Decis Mak 2015 Oct 14;15:86 [FREE Full text] [doi: 10.1186/s12911-015-0202-2] [Medline: 26466787]
6. Audrey S, Brown L, Campbell R, Boyd A, Macleod J. Young people's views about consenting to data linkage: findings from the PEARL qualitative study. BMC Med Res Methodol 2016 Mar 21;16:34 [FREE Full text] [doi: 10.1186/s12874-016-0132-4] [Medline: 27001504]
7. Haddow G, Bruce A, Sathanandam S, Wyatt JC. 'Nothing is really safe': a focus group study on the processes of anonymizing and sharing of health data for research purposes. J Eval Clin Pract 2011 Dec;17(6):1140-1146. [doi: 10.1111/j.1365-2753.2010.01488.x] [Medline: 20629997]
8. Hunn A. Survey of the General Public: Attitudes Towards Health Research. 2017. URL: https://s3.eu-west-2.amazonaws.com/www.hra.nhs.uk/media/documents/HRA_NIHR_general_public_omnibus_survey_2017_FINAL.pdf [accessed 2021-02-10]
9. Health Data Research UK. Pioneering data research centres to enable cutting-edge research and innovation to benefit UK patients. 2019 Sep 12. URL: https://www.hdruk.ac.uk/news/pioneering-data-research-centres-to-enable-cutting-edge-research-and-innovation/ [accessed 2021-02-10]
10. Miranda ML, Ferranti J, Strauss B, Neelon B, Califf RM. Geographic Health Information Systems: A Platform To Support The 'Triple Aim'. Health Affairs 2013 Sep;32(9):1608-1615. [doi: 10.1377/hlthaff.2012.1199]
11. Ahn SJ. Incorporating immersive virtual environments in health promotion campaigns: a construal level theory approach. Health Commun 2015;30(6):545-556. [doi: 10.1080/10410236.2013.869650] [Medline: 24991725]
12. Fraccaro P. The Nest-an interactive drama exploring the sharing of health data and patient records.: Greater Manchester Patient Safety Translational Research Centre; 2015. URL: https://gmpstrc.wordpress.com/2015/12/15/the-nest-an-interactive-drama-exploring-the-sharing-of-health-data-and-patient-records/ [accessed 2021-02-10]
13. Baum S. PatientsLikeMe wants to convince patients to share health data through social media campaign. MedCity News. 2014. URL: https://medcitynews.com/2014/03/patientslikeme-social-media-campaign-hopes-encourage-patients-share-data/ [accessed 2021-02-10]
14. Grossman LV, Masterson Creber RM, Ryan B, Restaino S, Alarcon I, Polubriaginof F, et al. Providers' perspectives on sharing health information through acute care patient portals. AMIA Annu Symp Proc 2018;2018:1273-1281 [FREE Full text] [Medline: 30815169]
15. Neves AL, Poovendran D, Freise L, Ghafur S, Flott K, Darzi A, et al. Health care professionals' perspectives on the secondary use of health records to improve quality and safety of care in England: qualitative study. J Med Internet Res 2019 Sep 26;21(9):e14135 [FREE Full text] [doi: 10.2196/14135] [Medline: 31573898]
16. Agaku IT, Adisa AO, Ayo-Yusuf OA, Connolly GN. Concern about security and privacy, and perceived control over collection and use of health information are related to withholding of health information from healthcare providers. J Am Med Inform Assoc 2014;21(2):374-378 [FREE Full text] [doi: 10.1136/amiajnl-2013-002079] [Medline: 23975624]
17. Dimitropoulos L, Patel V, Scheffler SA, Posnack S. Public attitudes toward health information exchange: perceived benefits and concerns. Am J Manag Care 2011 Dec;17(12 Spec No):SP111-SP116 [FREE Full text] [Medline: 22216769]
18. Deyo D, Khaliq A, Mitchell D, Hughes DR. Electronic sharing of diagnostic information and patient outcomes. Am J Manag Care 2018 Jan;24(1):32-37 [FREE Full text] [Medline: 29350510]
19. Bailey LC, Milov DE, Kelleher K, Kahn MG, Del Beccaro M, Yu F, et al. Multi-Institutional Sharing of Electronic Health Record Data to Assess Childhood Obesity. PLoS One 2013;8(6):e66192 [FREE Full text] [doi: 10.1371/journal.pone.0066192] [Medline: 23823186]
20. Neves AL, Poovendran D, Freise L, Ghafur S, Flott K, Darzi A, et al. Health care professionals' perspectives on the secondary use of health records to improve quality and safety of care in England: qualitative study. J Med Internet Res 2019 Sep 26;21(9):e14135 [FREE Full text] [doi: 10.2196/14135] [Medline: 31573898]

21. The One-way Mirror: Public Attitudes to Commercial Access to Health Data Internet. Wellcome Trust. 2016. URL: https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf [accessed 2021-02-10]
22. Data Sharing and Technology: Exploring the Attitudes of People With Asthma. Asthma UK. 2018. URL: https://tinyurl.com/2galaxca [accessed 2021-02-10]
23. Day S, Zweig M. Beyond Wellness For the Healthy: Digital Health Consumer Adoption 2018. Rock Health. 2018. URL: https://rockhealth.com/reports/beyond-wellness-for-the-healthy-digital-health-consumer-adoption-2018/ [accessed 2021-02-10]
24. Jagsi R, Griffith KA, Jones RD, Krenz C, Gornick M, Spence R, et al. Effect of public deliberation on patient attitudes regarding consent and data use in a learning health care system for oncology. J Clin Oncol 2019 Dec 1;37(34):3203-3211. [doi: 10.1200/JCO.19.01693] [Medline: 31577472]
25. Spagnuelo D, Lenzini G. Patient-centred transparency requirements for medical data sharing systems. In: Rocha A, Correia AM, Adeli H, Reis LP, Mendonça Teixeira M, editors. New Advances in Information Systems and Technologies. Berlin: Springer; 2016:1073-1083.
26. Rolfe DE, Ramsden VR, Banner D, Graham ID. Using qualitative Health Research methods to improve patient and public involvement and engagement in research. Res Involv Engagem 2018 Dec 13;4(1):-. [doi: 10.1186/s40900-018-0129-8]
27. Data Stories. Data Can of Worms. URL: https://datacanofworms.co.uk/data-stories [accessed 2021-02-16]
28. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
29. NIHR INVOLVE. Public involvement in research and research ethics committee review. 2016. URL: https://www.invo.org.uk/wp-content/uploads/2016/05/HRA-INVOLVE-updated-statement-2016.pdf [accessed 2021-02-10]
30. The New Economics Foundation. Who sees what: Exploring public views on personal electronic health records. 2010. URL: https://neweconomics.org/uploads/files/2cb17ab59382fe7c67_bfm6bdoas.pdf [accessed 2021-02-10]
31. Neves AL, Freise L, Laranjo L, Carter AW, Darzi A, Mayer E. Impact of providing patients access to electronic health records on quality and safety of care: a systematic review and meta-analysis. BMJ Qual Saf 2020 Jun 12:-. [doi: 10.1136/bmjqs-2019-010581]
32. Kim E, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Challenges to using an electronic personal health record by a low-income elderly population. J Med Internet Res 2009;11(4):e44 [FREE Full text] [doi: 10.2196/jmir.1256] [Medline: 19861298]
33. Woods SS, Forsberg CW, Schwartz EC, Nazi KM, Hibbard JH, Houston TK, et al. The association of patient factors, digital access, and online behavior on sustained patient portal use: a prospective cohort of enrolled users. J Med Internet Res 2017 Oct 17;19(10):e345 [FREE Full text] [doi: 10.2196/jmir.7895] [Medline: 29042345]
34. Shull JG. Digital health and the state of interoperable electronic health records. JMIR Med Inform 2019 Dec 1;7(4):e12712 [FREE Full text] [doi: 10.2196/12712] [Medline: 31682583]
35. Aitken M, Tully M, Porteous C, Denegri S, Cunningham-Burley S, Banner N, et al. Consensus statement on public involvement and engagement with data-intensive health research. IJPDS 2019 Feb 12;4(1):n/a-n/a [FREE Full text] [doi: 10.23889/ijpds.v4i1.586]
36. O'Brien EC, Rodriguez AM, Kum H, Schanberg LE, Fitz-Randolph M, O'Brien SM, et al. Corrigendum to "Patient perspectives on the linkage of health data for research: Insights from an online patient community questionnaire" [Int. J. Med. Inform. 127 (2019) 9–17]. Int J Med Info 2020 Apr;136:104084. [doi: 10.1016/j.ijmedinf.2020.104084]
37. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Affairs 2014 Jul 8;33(7):1123-1131. [doi: 10.1377/hlthaff.2014.0041] [Medline: 25006137]
38. John A, Marchant A, McGregor J, Tan J, Hutchings H, Kovess V, et al. Recent trends in the incidence of anxiety and prescription of anxiolytics and hypnotics in children and young people: An e-cohort study. J Affective Disorders 2015 Sep;183:134-141. [doi: 10.1016/j.jad.2015.05.002]
39. Burgess C, Kauth MR, Klemt C, Shanawani H, Shipherd JC. Evolving sex and gender in electronic health records. Fed Pract 2019 Jul;36(6):271-277 [FREE Full text] [Medline: 31258320]
40. Public Attitudes to Patient Data Use: a Summary of Existing Research Internet. Understanding Patient Data. 2018. URL: https://understandingpatientdata.org.uk/sites/default/files/2018-08/Public [accessed 2021-02-10]
41. Neves AL, Poovendran D, Freise L, Ghafur S, Flott K, Darzi A, et al. Health care professionals' perspectives on the secondary use of health records to improve quality and safety of care in England: qualitative study. J Med Internet Res 2019 Sep 26;21(9):e14135 [FREE Full text] [doi: 10.2196/14135] [Medline: 31573898]
42. Aitken M, de SJJ, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016 Nov 10;17(1):73 [FREE Full text] [doi: 10.1186/s12910-016-0153-x] [Medline: 27832780]
43. Wilson A. Health care professionals' perspectives on the secondary use of health records to improve quality and safety of care in england: qualitative study. Improving Data Quality to Deliver Better Patient Outcomes. 2019 Jan 15. URL: https://thejournalofmhealth.com/improving-data-quality-to-deliver-better-patient-outcomes/ [accessed 2021-02-10]

44. Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data internet. Wellcome Trust. 2013. URL: https://wellcome.ac.uk/sites/default/files/wtp053205_0.pdf [accessed 2021-02-10]
45. Weiner M, Biondich P. The influence of information technology on patient-physician relationships. J Gen Intern Med 2006 Jan;21(Suppl 1):S35-S39 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00307.x] [Medline: 16405708]
46. Al-Busaidi ZQ. Qualitative research and its uses in health care. Sultan Qaboos Univ Med J 2008;8(1):11-19 [FREE Full text] [Medline: 21654952]
47. Suter W. Qualitative data, analysis, and design. In: Introduction to Educational Research: A Critical Thinking Approach. Thousand Oaks: SAGE Publications; 2012.
48. Booth A, Hannes K, Harden A, Noyes J, Harris J, Tong A. COREQ (Consolidated Criteria for Reporting Qualitative Studies): Guidelines for Reporting Health Research: A User's Manual. New Jersey, US: John Wiley & Sons; 2014.
49. Information Commissioner's Office. Trust and confidence in data. 2017. URL: https://2sjjwunnql41ia7ki31qqub1-wpengine.netdna-ssl.com/wp-content/uploads/2017/11/ICO_trust-and-confidence-in-data_2017-1.pdf [accessed 2021-02-10]

Abbreviations

EHR: electronic health record
HRA: Health Research Authority
NHS: National Health Service

Edited by G Eysenbach; submitted 28.07.20; peer-reviewed by J Shull, J van Delden, KL Ong; comments to author 26.10.20; revised version received 27.11.20; accepted 16.01.21; published 22.02.21.

Please cite as:
Lounsbury O, Roberts L, Goodman JR, Batey P, Naar L, Flott KM, Lawrence-Jones A, Ghafur S, Darzi A, Neves AL
Opening a "Can of Worms" to Explore the Public's Hopes and Fears About Health Care Data Sharing: Qualitative Study
J Med Internet Res 2021;23(2):e22744
URL: https://www.jmir.org/2021/2/e22744
doi:10.2196/22744
PMID:33616532

Viewpoint

Revolutionizing Medical Data Sharing Using Advanced Privacy-Enhancing Technologies: Technical, Legal, and Ethical Synthesis

James Scheibner[1,2], BComp, LLB, PhD; Jean Louis Raisaro[3,4], BSc, MSc, PhD; Juan Ramón Troncoso-Pastoriza[5], BSc, MSc, MPhil, PhD; Marcello Ienca[1], BA, MA, MSc, PhD; Jacques Fellay[3,6,7], MD, PhD; Effy Vayena[1], BA, MSc, PhD; Jean-Pierre Hubaux[5], Dr-Eng

[1]Health Ethics and Policy Laboratory, Department of Health Sciences and Technology, Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland

[2]College of Business, Government and Law, Flinders University, Adelaide, Australia

[3]Precision Medicine Unit, Lausanne University Hospital, Lausanne, Switzerland

[4]Data Science Group, Lausanne University Hospital, Lausanne, Switzerland

[5]Laboratory for Data Security, School of Computer and Communication Sciences, École polytechnique fédérale de Lausanne, Lausanne, Switzerland

[6]School of Life Sciences, École polytechnique fédérale de Lausanne, Lausanne, Switzerland

[7]Host-Pathogen Genomics Laboratory, Swiss Institute of Bioinformatics, Lausanne, Switzerland

Corresponding Author:
James Scheibner, BComp, LLB, PhD
College of Business, Government and Law
Flinders University
Ring Road, Bedford Park
Adelaide, 5042
Australia
Phone: 61 (08) 8201 3196
Email: james.scheibner@flinders.edu.au

Abstract

Multisite medical data sharing is critical in modern clinical practice and medical research. The challenge is to conduct data sharing that preserves individual privacy and data utility. The shortcomings of traditional privacy-enhancing technologies mean that institutions rely upon bespoke data sharing contracts. The lengthy process and administration induced by these contracts increases the inefficiency of data sharing and may disincentivize important clinical treatment and medical research. This paper provides a synthesis between 2 novel advanced privacy-enhancing technologies—homomorphic encryption and secure multiparty computation (defined together as multiparty homomorphic encryption). These privacy-enhancing technologies provide a mathematical guarantee of privacy, with multiparty homomorphic encryption providing a performance advantage over separately using homomorphic encryption or secure multiparty computation. We argue multiparty homomorphic encryption fulfills legal requirements for medical data sharing under the European Union's General Data Protection Regulation which has set a global benchmark for data protection. Specifically, the data processed and shared using multiparty homomorphic encryption can be considered anonymized data. We explain how multiparty homomorphic encryption can reduce the reliance upon customized contractual measures between institutions. The proposed approach can accelerate the pace of medical research while offering additional incentives for health care and research institutes to employ common data interoperability standards.

(J Med Internet Res 2021;23(2):e25120) doi:10.2196/25120

KEYWORDS

encryption; anonymization; pseudonymization; centralized approach; decentralized approach; federated approach; Interoperability; privacy; GDPR; General Data Protection Regulation; data privacy; data protection; ethics; research; data sharing; data governance; patient data privacy

Introduction

The current biomedical research paradigm has been characterized by a shift from intrainstitutional research toward multiple collaborating institutions operating at an interinstitutional, national or international level for multisite research projects; however, despite the apparent breakdown of research barriers, there remain differences between ethical and legal requirements at all jurisdictional levels [1]. There are numerous organizational strategies that have been used to resolve these issues, particularly for international academic consortia.

For example, the International Cancer Genome Consortium endeavors to amass cancer genomes paired with noncancerous sequences in a cloud environment, known as pancancer analysis of whole genomes. The International Cancer Genome Consortium's data access compliance office was unable to establish an international cloud under the Pancancer Analysis of Whole Genomes Project because of conflicts between United States and European Union data privacy laws [2]. These conflicts will be likely exacerbated with the Court of Justice of the European Union (CJEU) invalidating the United States–European Union Privacy Shield agreement. This decision will prevent private research organizations from transferring personal data from the European Union to the United States without organizational safeguards [3]. In addition, the COVID-19 pandemic has made sharing data for clinical trials and research imperative. However, a series of COVID-19 papers retracted due to data unavailability emphasizes the need for data sharing to encourage oversight [4]. Furthermore, within the European Union there is the potential for differences in how countries regulate the processing of health-related personal data [5]. There are also different grounds to justify processing of health-related data under separate branches of EU law. The Clinical Trials Regulation and the European Union General Data Protection Regulation (GDPR) require different standards of consent for processing health-related data, depending on whether those data are collected as part of a clinical trial protocol or not. The effect of this difference is that data collected for one purpose, such as a trial protocol, may not be made available for a secondary research purpose if appropriate consent has not been obtained [6]. Finally, given study restrictions it may be impossible to share data between institutions or jurisdictions [7]. Although reforms to EU data protection law have been proposed to encourage scientific data sharing [8], at present the best available solutions remain contractual and technological measures.

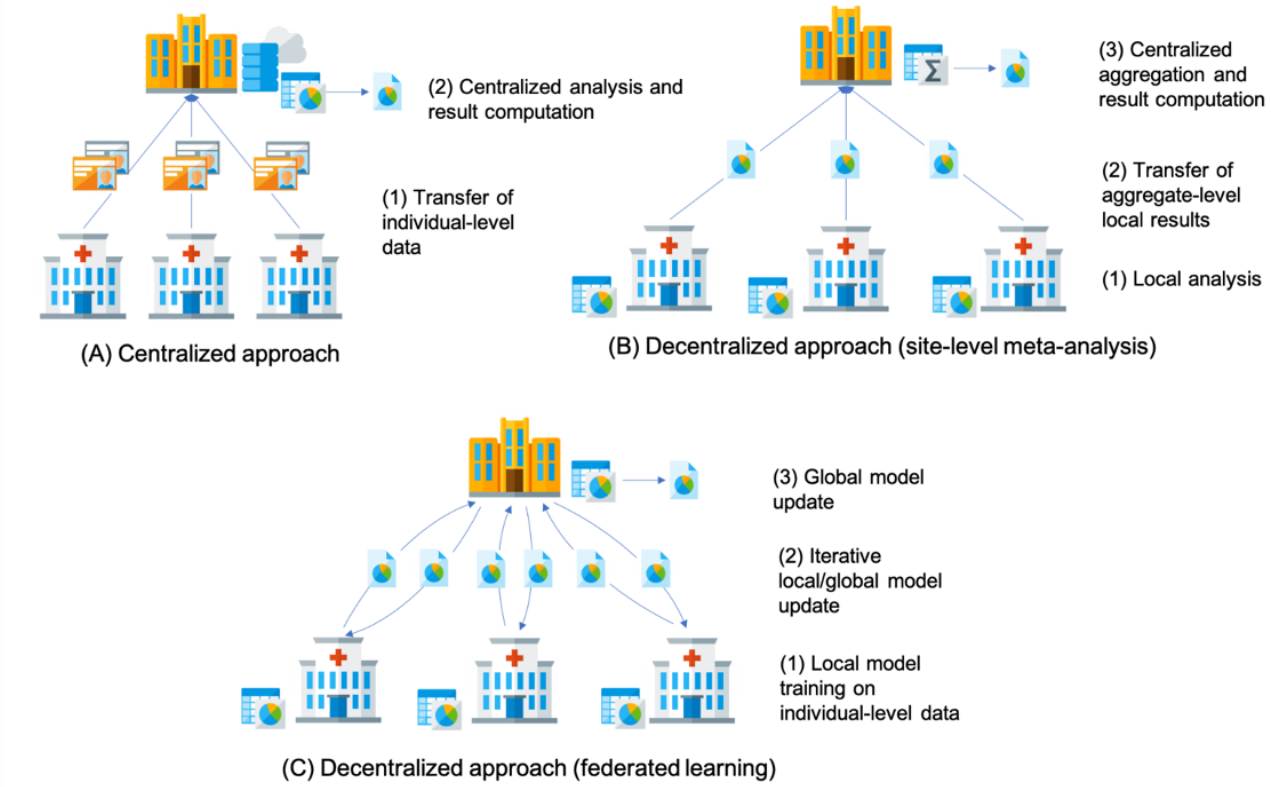
In this paper, we describe how traditional data-sharing approaches relying upon conventional privacy-enhancing technologies are limited by various regulations governing medical use and data sharing. We describe two novel privacy-enhancing technologies, homomorphic encryption and secure multiparty computation, that extend the capacity of researchers to conduct privacy-preserving multisite research. We then turn to analyze the effects of regulation on using these novel privacy-enhancing technologies for medical and research data sharing. In particular, we argue these privacy-enhancing technologies guarantee anonymity as defined under the EU GDPR and are, therefore, key enablers for medical data sharing. We focus on the GDPR, as it currently represents a global benchmark in data protection regulations. We argue that using these technologies can reduce the reliance upon customized data-sharing contracts. The use of standardized agreements for multiparty processing of data in concert with privacy-enhancing technologies can reduce the bottleneck on research. Finally, we turn to address how these novel privacy-enhancing technologies can be integrated within existing regulatory frameworks to encourage increased data sharing while preserving data privacy.

Privacy and Security Issues of Current Medical Data-Sharing Models

Overview

Before examining novel privacy-enhancing technologies, it is necessary to examine the main models for exchanging medical data for research purposes and the limitations of conventional privacy protection mechanisms that are currently used to reduce the risk of reidentification. We synthesize the data-sharing models into three categories and analyze their main technological issues (Figure 1).

Figure 1. Overview of the three main data-sharing models: (A) centralized, (B) decentralized (site-level meta-analysis), and (C) decentralized (federated learning).



Centralized Model: Trusted Dealer

The centralized model requires medical sites (ie, data providers) that are willing to share data with each other to pool their individual-level patient data into a single repository. The data repository is usually hosted by one medical site or by an external third party (eg, a cloud provider), playing the trusted dealer role. The main advantage of this model is that the trusted dealer enables authorized investigators to access all the patient-level information needed for data cleaning and for conducting statistical analysis. Moreover, such a data-sharing model minimizes infrastructure costs at medical sites, as data storage and computation are outsourced. However, from a data privacy perspective the centralized model is often difficult to realize, especially when medical and genetic data should be exchanged across different jurisdictions. The central site hosting the data repository represents a single point of failure in the data-sharing process. All participating sites must trust such single entity for protecting their patient-level data [9].

To minimize sensitive information leakage from data breaches, traditional anonymization techniques include suppressing directly identifying attributes, as well as the generalizing, aggregating or randomizing quasi-identifying attributes in individual patient records. In particular, the *k*-anonymity privacy model [10] is a well-established privacy-preserving model that aims to reduce the likelihood of reidentification attacks singling out an individual. Specifically, the *k*-anonymity model ensures that for each combination of quasi (or indirect) identifier, there exists at least *k* individuals who share the same attributes.

However, given the increased sophistication of reidentification attacks [10-16] and the rising dimensionality (number of clinical and genetic attributes) of patient data, the above-mentioned countermeasures are inadequate to ensure a proper level of anonymization and preserve acceptable data utility. As a result, these conventional anonymization techniques for individual-level patient data are rarely used in practice. Researchers prefer to rely upon simple pseudonymization techniques (such as replacing direct identifiers with pseudonymous codes) combined with legal measures defining each party's responsibilities regarding data transfer, access, and use. This process generates administrative overheads that slow down the pace of biomedical research. Furthermore, although designed to comply with data protection regulations, contractual safeguards may not eliminate the risk of individuals being reidentified [17]. As we argue below, combining traditional pseudonymization mechanisms and governance strategies meets the legal standard of pseudonymization but not anonymization under the GDPR.

Decentralized Model: Site-Level Meta-analysis

As opposed to the centralized data-sharing model, the decentralized model does not require patient-level data to be physically transferred out of the medical sites' information technology infrastructure. Medical sites keep control over their individual-level patient data and define their own data governance principles. For each clinical study, the statistical analysis is first computed on local data sets. The resulting local statistics are then sent to the site responsible for the final meta-analysis that aggregates the separate contribution of each data provider [18] to obtain the final result of the analysis. Under

this model, the site performing the meta-analysis is trusted by all other sites for the protection of their local statistics. As local statistics have a significantly lower dimensionality with respect to individual-level data, there is a lower risk of reidentification in the decentralized data-sharing model.

However, the sharing of only aggregate-level data does not guarantee patients' privacy by itself. Some aggregate-level statistics may be too low for certain subpopulations (such as patients with rare diseases) and can be considered personally identifying. Moreover, in some circumstances aggregate-level data from local analyses can be exploited to detect the presence of target individuals in the original data set. For example, an attacker may already hold the individual-level data of 1 or several target individuals [19-23]. This membership information can be subsequently used to infer sensitive and sometimes stigmatizing attributes of the target individuals. For example, detecting the membership of an individual in a HIV-positive cohort reveals their HIV status. The intuition behind these attacks is to measure the similarity between the individual-level target data with statistics computed from the study data set and statistics computed from the general population. The attacker's certainty about the target's membership in the data set increases with the similarity of the target's data to the statistics derived from the study data set.

To address these inference attacks, clinical sites can anonymize their local statistics by applying obfuscation techniques that mainly consist in adding a certain amount of statistical noise on the aggregate-level data before transfer to third parties. This process enables data providers to achieve formal notions of privacy such as differential privacy [24,25]. In the statistical privacy community, differential privacy is currently considered as guaranteeing the likelihood of reidentification from the release of aggregate-level statistics can be minimized to an acceptable value. Similar to anonymization techniques for individual-level data, statistical obfuscation techniques degrade the utility of aggregate-level data. Consequently, the amount of noise introduced by data obfuscation should be carefully calibrated to reach the desired compromise between utility and privacy. Often, when each data provider adds the required amount of noise to reach an acceptable level of privacy, the resulting aggregated results stemming from a meta-analysis are too distorted to be reliable [26].

Beyond privacy considerations, this approach also suffers from a lack of flexibility as the medical sites involved in the analysis must coordinate before the analysis on the choice of parameters and covariates to be considered. This coordination often depends on manual approval, impeding the pace of the analysis itself. Finally, as opposed to the centralized approach, accuracy of results from a meta-analysis that combines the summary statistics or results of local analysis can be affected by cross-study heterogeneity. This can lead to inaccurate and misleading conclusions [27].

Decentralized Model: Federated Analysis and Learning

The federated model is an evolution of the decentralized model based on site-level meta-analysis. Instead of sharing the results of local analyses, the participating data providers collaborate to perform a joint analysis or the training of a machine learning model in an interactive and iterative manner, only sharing updates of the model's parameters. One of the medical sites participating in the multicentric research project (typically the site responsible for the statistical analysis) becomes the reference site (or central site) and defines the model to be trained (or analysis to be performed) and executed on the data distributed across the network. This model is referred to as the global model. Each participating site is given a copy of the model to train on their own individual-level data. Once the model has been trained locally over several iterations, the sites send only their updated version of the model parameters (aggregate-level information) to the central site and keep their individual-level data at their premises. The central site aggregates the contributions from all the sites and updates the global model [28]. Finally, the updated parameters of the global model are shared again with the other sites. The process repeats iteratively till convergence of the global model.

With respect to the distributed data-sharing approach based on site-level meta-analysis, this federated approach is more robust against heterogeneous distributions of the data across different sites, thus yielding results accuracy that is comparable to the results obtained with the same analysis conducted using the centralized model. Moreover, this approach does not suffer from the loss in statistical power of conventional meta-analyses. Prominent projects that have attempted to employ federated approaches to analysis and sharing of biomedical data are the DataSHIELD project [29] and the Medical Informatics Platform of the Human Brain Project [30].

The federated data-sharing approach combines the best features of the other two approaches. However, although the risk or reidentification is reduced compared to the centralized approach, the federated approach remains vulnerable to the same inference attacks of the meta-analysis approach. These inference attacks exploit aggregate-level data released during collaboration [31-34]. The potential for an inference attack is even increased compared to a meta-analysis-based approach. This is due to the iterative and collaborative nature of the data processing, allowing adversaries to observe model changes over time and with specific model updates. Melis et al [35] show that updates of model parameters transferred during the collaborative training phase can be used to infer the membership of a target individual in the training data sets as well as some properties associated with a particular subset of the training data. This inference is possible if the context of the data release enables the attacker to easily access some auxiliary individual-level information about the target individual. In legal terms (as discussed below), these aggregate-level data can potentially be considered personal data. As for the meta-analysis approach, obfuscation techniques can be used to anonymize the model's updates at each iteration. Nevertheless, the required perturbation can severely affect the performance of the final model [26].

Finally, regardless of the type of distributed data-sharing model, obfuscation techniques for anonymizing aggregate-level data are rarely used in practice in medical research because of their impact on data utility. As a result, these technical privacy limitations are usually addressed via additional legal and organizational mechanisms. For the DataSHIELD project, access is limited to organizations that have consented to the terms of

use for DataSHIELD and have sought appropriate ethics approval to participate in a DataSHIELD analysis [36]. Therefore, implementing the platform will require cooperating with governments and institutions so they are comfortable with exposing sensitive data to the platform [29]. However, as we discuss below, advanced technologies can also guarantee data privacy.

Minimizing Risks by Leveraging Advanced Privacy-Enhancing Technologies

Overview

In the last few years, several cryptographic privacy-enhancing technologies have emerged as significant potential advances for addressing the above-mentioned data protection challenges that still affect medical data sharing in the decentralized model. Although hardware-based approaches could be envisioned for this purpose, they are usually tailored to centralized scenarios and introduce a different trust model involving the hardware provider. Furthermore, they also depend on the validity of the assumptions on the security of the hardware platform, for which new vulnerabilities are constantly being discovered. In this paper, we focus on two of the most powerful software-based privacy-enhancing technologies: homomorphic encryption and secure multiparty computation. Both rely upon mathematically proven guarantees for data confidentiality, respectively grounded on cryptographic hard problems and noncollusion assumptions.

Homomorphic Encryption

Homomorphic encryption [37] is a special type of encryption that supports computation on encrypted data (ciphertexts) without decryption. Thanks to this property, homomorphically encrypted data can be securely handed out to third parties, who can perform meaningful operations on them without learning anything about their content. Fully homomorphic encryption schemes, or schemes enabling arbitrary computations on ciphertexts, are still considered nonviable due to the high computational and storage overheads they introduce. Current practical schemes that enable only a limited number of computations on ciphertexts (such as polynomial operations) have reached a level of maturity that permits their use in real scenarios.

Secure Multiparty Computation

Secure multiparty computation [38-42] protocols enable multiple parties to jointly compute functions over their private inputs without disclosing to the other parties more information about each other's inputs than what can be inferred from the output of the computation. This class of protocols is particularly attractive in privacy-preserving distributed analytic platforms due to the great variety of secure computations they enable. However, this flexibility includes several drawbacks that hinder their adoption, including high network overhead and the requirement for parties to remain online during computation.

Multiparty Homomorphic Encryption

The combination of secure multiparty computation and homomorphic encryption was proposed to overcome their respective overheads and technical limitations; we refer to it as multiparty homomorphic encryption [43-46]. Multiparty homomorphic encryption enables flexible secure processing by efficiently transitioning between encrypted local computation, performed with homomorphic encryption, and interactive protocols (secure multiparty computation). It can be used to choose the most efficient approach for each step within a given workflow, leveraging the properties of one technique to avoid the bottlenecks of the other. Moreover, multiparty homomorphic encryption ensures that the secret key of the underlying homomorphic encryption scheme never exists in full. Instead, it distributes the control over the decryption process across all participating sites, each one holding a fragment of the key. All participating sites have to agree to enable the decryption of any piece of data, and no single entity alone can decrypt the data.

Unlike homomorphic encryption or secure multiparty computation alone, multiparty homomorphic encryption provides effective, scalable, and practical solutions for addressing the privacy-preserving issues that affect the distributed or federated approach for data sharing. For example, systems such as Helen [47], MedCo [48], or POSEIDON [49] use multiparty homomorphic encryption to guarantee that all the information interchanged between the sites is always in encrypted form, including aggregate data such as model parameters and model updates, and only the final result (the computed model or the predictions based on this model) is revealed to the authorized user. Finally, multiparty homomorphic encryption reduces the need of obfuscation techniques to protect aggregate-level data from inference attacks. Furthermore, data utility, which is typically lost with privacy-preserving distributed approaches that only rely upon obfuscation techniques, can be significantly improved. As aggregate-level data transfer and processing across participating sites during the analysis or training phase remains always encrypted, obfuscation can be applied only to the decrypted final result of the analysis that is released to the data analyst, instead of being applied to all local model updates at each iteration. Hence, multiparty homomorphic encryption enables a much lower utility degradation for the same level of reidentification risk.

Regulatory Hurdles for the Use of Encryption Technologies

Overview

In this section, we focus on the features of EU data protection law concerning encryption and data sharing. We focus on the GDPR because of the persistence of national divergences in member state law, despite the passage of the GDPR. In particular, the GDPR provides member states can introduce further conditions, including restrictions on processing of genetic data, biometric data, or health-related data. These exceptions exist outside the narrow circumstances in which special categories of personal data, which genetic data, biometric data, or health-related data belong to, can be processed [6]. This flexibility increases the potential for divergences in national law that require customized contracts between institutions in different EU member states [5].

Data Anonymization and Pseudonymization

The GDPR defines *personal data* as concerning an identifiable natural person. Therefore, pseudonymized data, where all identifiers have been removed from those data, remain personal data. However, the provisions of the GDPR do not concern anonymized data or data which have been processed so individuals are no longer identifiable. In particular, anonymized data may be used for research or statistical processing without the need to comply with the GDPR.

Spindler and Schmechel [50] note there are two conflicting approaches to classifying personal and anonymized data. The first is an absolute approach, where anonymized data constitute personal data if there is even a theoretical chance of reidentification. This approach represents the state of national law in a minority of EU member states, such as France [51]. The second is the relative approach, where anonymized data are no longer personal data if it is reasonably likely that methods do not exist to reidentify individuals [50]. This approach represents the state of national law in countries such as Ireland, where the Irish Data Protection Commission has held that data are anonymized if it is unlikely current technology can be used to reidentify those data [52]. Likewise, the German Federal Ministry for Economic Affairs and Energy held that data (including health-related personal data) are anonymized under the *Bundesdatenschutzgesetz* (*German Federal Data Protection Act*) where individuals cannot be reidentified with reasonable effort [53]. In both these jurisdictions, if an unreasonable effort were required to reidentify anonymized data, then it would no longer be personal data [50].

At the supranational level, the former Article 29 Working Party (now the European Data Protection Board) has favored a relative over an absolute approach to anonymization. First, the Article 29 Working Party held that the words "means reasonably likely" suggests a theoretical possibility of reidentification will not be enough to render those data personal data [54]. A subsequent opinion of the Working Party reinforced this support for the relative approach and compared different techniques for anonymization or pseudonymization. For example, encrypting data with a secret key means that data could be decrypted by the key holder. For this party, the data would therefore be pseudonymized data. But, if a party does not have the key, the data would be anonymized. Likewise, if data are aggregated to a sufficiently high level, these data would no longer be personal data [55]. Nevertheless, following the Article 29 Working Party's ruling, no single anonymization technique can fully guard against orthogonal risks of reidentification [56].

Data Processing

The GDPR's provisions apply to data controllers, or entities determining the purpose and means of processing personal data. This definition encompasses both health care institutions and research institutions. Data controllers must guarantee personal data processing is lawful, proportionate, and protects the rights of data subjects. In particular, the GDPR provides that encryption should be used as a safeguard when personal data are processed for a purpose other than which they were collected. Although the GDPR does not define encryption, the Article 29 Working Party treats encryption as equivalent to stripping identifiers from personal data. The GDPR also lists encryption as a strategy that can guarantee personal data security. Furthermore, the GDPR emphasizes that data controllers should consider the state of the art, along with the risks associated with processing, when adopting security measures. The GDPR also provides that data processing for scientific purposes should follow the principle of data minimization. This principle requires data processors and controllers to use nonpersonal data unless the research can only be completed with personal data. If personal data are required to complete the research, pseudonymized or aggregate data should be used instead of directly identifying data.

The GDPR imposes obligations on data controllers with respect to the transfer of data, particularly outside of the European Union. Specifically, the GDPR requires the recipient jurisdiction to offer adequate privacy protection before a data controller transfers data there. Otherwise, the data controller must ensure there are organizational safeguards in place to ensure the data receives GDPR-equivalent protection. Furthermore, data controllers must consider the consequences of exchanging data between institutions, and whether these are joint controllership or controller–processor arrangements. Under the GDPR, data subject rights can be exercised against any and each controller in a joint controllership agreement. Furthermore, controllers must have in place an agreement setting out the terms of processing. By contrast, a data controller-processor relationship exists where a controller directs a data processor to perform processing on behalf of the controller, such as a cloud services provider. The GDPR provides that any processing contract must define the subject matter, duration, and purpose of processing. Contracts should also define the types of personal data processed and require processors to guarantee both the confidentiality and security of processing.

Advanced Privacy-Enhancing Technologies and EU Data Governance Requirements

In this section, we argue that multiparty homomorphic encryption, or homomorphic encryption and secure multiparty computation used in concert, meets the requirements for anonymization of data under the GDPR. Furthermore, we argue the use of multiparty homomorphic encryption can significantly reduce the need for custom contracts to govern data sharing between institutions. We focus on genetic and clinical data sharing due to the potential for national derogations pertaining to the processing of health-related data. Nevertheless, our conclusions regarding the technical and legal requirements for data sharing using multiparty homomorphic encryption, or homomorphic encryption and secure multiparty computation, may apply to other sectors, depending on regulatory requirements [57].

Under the GDPR, separating pseudonymized data and identifiers is analogous to separating decryption keys and encrypted data. For pseudonymized data, any entity with physical or legal access to the identifiers will possess personal data [58]. To this end, Spindler and Schmechel [50] suggest that encrypted data remain personal data to the entity holding the decryption keys. The encrypted data also remain personal data for any third party with lawful means to access the decryption keys. Applying this

approach to homomorphic encryption, if a party has access to the decryption key corresponding to the encryption key that was used to homomorphically encrypt data, that party will have access to personal data. Likewise, if a party has lawful access to data jointly processed as part of secure multiparty computation, those data will remain personal data for that party [59].

Whether a party to data processing using advanced privacy-enhancing technologies has lawful access to data or decryption keys depends on the legal relationship between the parties. With respect to joint controllership, recent CJEU case law has established that parties can be joint controllers even without access to personal data [60-62]. The CJEU held that the administrator of a fan page hosted on Facebook was a joint controller despite only having access to aggregate data in paragraph 38 [60];however, Article 26, paragraph 1 of the GDPR [63] requires that joint controllers establish a contract allocating responsibility for processing of personal data. Hospitals or research institutions processing patient data using secure multiparty computation jointly determine how these data are processed. These entities would be classified as joint controllers, at least when engaging in secret sharing (as a joint purpose of data processing). These entities would need an agreement to establish that only the entity with physical access to patient data can access those data. If a request is made to a hospital or research institution that does not possess these data, the request must be referred to the entity that does.

Applying these principles to processing with privacy-enhancing technologies, for homomorphic encryption, there is no mathematical possibility of decrypting the data without the decryption key. This holds true when both the data are at rest or when the data are processed in the encrypted space via secure operations such as homomorphic addition or multiplication. Whether data processed as part of secure multiparty computation or multiparty homomorphic encryption remain personal data depends on whether entities have lawful access to personal data or decryption keys respectively. If entities can only access personal data they physically hold as part of a joint controller agreement, the data fragments exchanged during secret sharing via secure multiparty computation are not personal data. Likewise, under multiparty homomorphic encryption each individual entity only has access to a fragment of the decryption key, which can only be recombined with the approval of all other entities holding the remaining fragments. This argument is reinforced by Recital 57 of the GDPR [63], which provides controllers forbidden from identifying individuals are not required to collect identifying information to comply with the GDPR.

Therefore, we submit that both homomorphic encryption and secure multiparty computation, when used alone or together through multiparty homomorphic encryption can jointly compute health-related data while complying with the GDPR. These data remain anonymous even though entities processing data using multiparty homomorphic encryption are joint controllers. Furthermore, the use of advanced privacy-enhancing technologies should become a best standard for the processing of health-related data for three reasons. First, the Article 29 Working Party has recommended using encryption and anonymization techniques in concert to protect against orthogonal privacy risks and overcome the limits of individual techniques [55]. Second, the GDPR emphasizes the use of state-of-the-art techniques for guaranteeing the processing of sensitive data. Homomorphic encryption, secure multiparty computation, and multiparty homomorphic encryption are considered state-of-the-art technologies in that they carry a mathematical guarantee of privacy. Third, the Article 29 Working Party has held the data controller is responsible for demonstrating that the data have been and remain anonymized [55]. Further support from this argument comes from a case heard before the Swiss Federal Supreme Court [64]; in paragraph 5.12, the Federal Supreme Court endorsed a relative approach to anonymization, but also placed the onus on the data controller to establish anonymization. Switzerland is not a member of the European Union and does not have to comply with the GDPR. However, Switzerland's close proximity to the European Union means the Swiss Federal Act on Data Protection has been revised. These revisions ensure the continued free exchange of data between Switzerland and EU countries [65].

Therefore, we argue that multiparty homomorphic encryption involves processing anonymized data under EU data protection law. Although homomorphic encryption, secure multiparty computation, and multiparty homomorphic encryption do not obliviate the need for a joint controllership agreement, they lessen the administrative burden required for data sharing. Furthermore, they promote the use of standard processing agreements that can help ameliorate the impacts of national differences within and outside the European Union. Accordingly, we submit that multiparty homomorphic encryption, along with other forms of advanced privacy-enhancing technologies, should represent the standard for health data processing in low trust environments [66]. This processing can include performing computations on sensitive forms of data, such as providing genomic diagnoses without revealing the entire sequence for a patient [67]. Furthermore, the encrypted outputs of homomorphic encryption and secure multiparty computation are mathematically private, as they do not reveal any personal data [68]. Finally, the fact that multiparty homomorphic encryption involves processing anonymized data broadens the purposes for which health-related data can be used. For example, as part of a clinical trial protocol data might be collected from patients via their personal devices. These devices can either store these data locally or transmit them to a hospital. The data may then be queried in an anonymized form as part of a research project without needing to seek additional consent that would otherwise be required under data protection law for those data to be processed [69]. The ability to reuse data that are stored on a patient's personal device can also help support innovative forms of clinical trials, such as remote patient monitoring. The various states of data processed using novel privacy-enhancing technologies such as multiparty homomorphic encryption is displayed in Figure 2. Table 1 demonstrates the status of personal data at different stages of processing.

Figure 2. Comparison of the status of personal data under a distributed approach relying upon traditional privacy-enhancing technologies (eg, aggregation and pseudonymization) and a distributed approach relying on multiparty homomorphic encryption (eg, homomorphic encryption and secure multiparty computation).

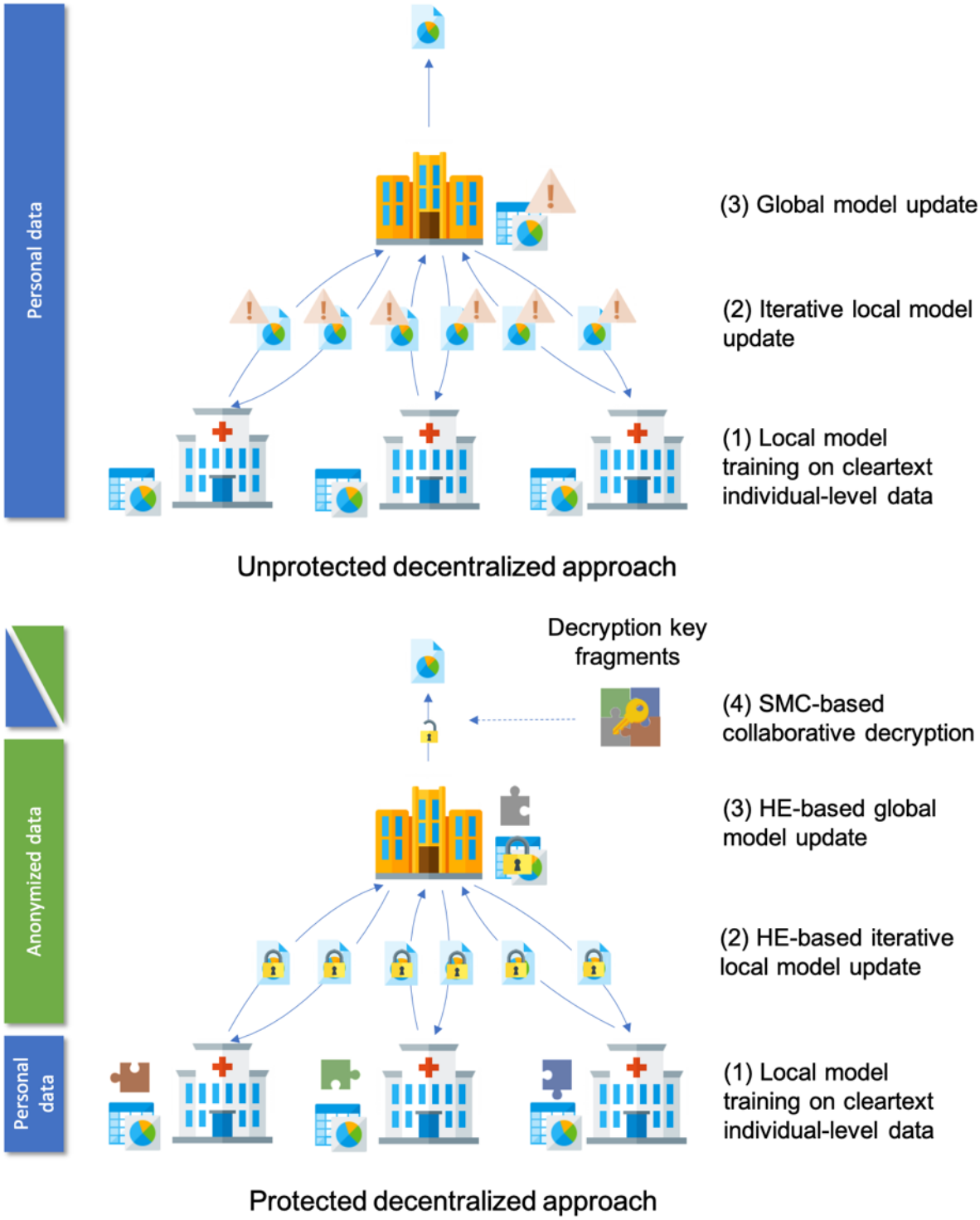


Table 1. Data status at different stages of processing.

Scenario	Description	Status of data based on the scenario
A	Hospital/research institution physically holds personal data	Personal data
B	Hospital/research institution has legal access to decryption key/personal data	Pseudonymized data
C	Hospital/research institution combine decryption keys/personal data to process data	Anonymized data
D	Third party (cloud service provider) carries out processing, hospitals share encryption keys jointly	Anonymized data

The lack of reliance upon custom contracts may encourage institutions to align their data formats to common international interoperability standards. In the next section, we turn to address the standardization of these advanced privacy-enhancing technologies.

Regulatory Instruments to Encourage the Use of Novel Privacy-Enhancing Technologies

At present, regulatory instruments provide limited guidance on the different types of privacy-enhancing technologies required to process medical data in a privacy-conscious fashion. However, the techniques described in this paper may represent a future best standard for processing medical data for clinical or research purposes. Because of the novelty of both technologies, the standardization of homomorphic encryption and secure multiparty computation is ongoing, with the first community standard released in 2018 [70].

Furthermore, there are numerous documents published by data protection agencies that can aid the development of such guidelines. For example, the *Commission Nationale de l'Informatique et des Libertés* (*French Data Protection Agency*) published a set of guidelines following the passage of the GDPR on how to secure personal data. This document provides recommendations on when encryption should be used, including for data transfer and storage [71]. Likewise, the *Agencia Española de Protección de Datos* (*Spanish Data Protection Agency*) has already recommended using homomorphic encryption as a mechanism for achieving data privacy by design pursuant to Article 25 of the GDPR [72].

Nevertheless, any standards will need to be continually updated to respond to new technological changes. For example, one of the most significant drawbacks of fully homomorphic encryption is the complexity of computation. This computational complexity makes it hard to predict running times, particularly for low-power devices such as wearables and smartphones. For the foreseeable future, this may limit the devices upon which fully homomorphic encryption can be used [73]. Therefore, specialized standards may need to be developed for using homomorphic encryption on low-power devices in a medical context. Specifically, these standards must be compliant with the legal requirements for access to and sharing of data by patients themselves, including the right to data portability as contained within Article 20 of the GDPR [**54**]. Although homomorphic encryption and secure multiparty computation offer privacy guarantees, there is still an orthogonal risk of reidentifying individuals from aggregate-level results that are eventually decrypted and can be exploited by inference attacks [19,21,27,74]. However, as mentioned earlier, the use of multiparty homomorphic encryption or secure multiparty computation enables the application of statistical obfuscation techniques for anonymizing aggregate-level results with a better privacy-utility trade-off than the traditional distributed approach, thus facilitating the implementation of end-to-end anonymized data workflows.

A final consideration relates to ethical issues that exist beyond whether homomorphic encryption, multiparty computation, and multiparty homomorphic encryption involve processing anonymized or personal data. First, the act of encrypting personal data constitutes further processing of those data under data protection law. Therefore, health care and research institutions must seek informed consent from patients or research participants [50]. Institutions must consider how to explain these technologies in a manner that is understandable and enables the patient to exercise their rights under data protection law. Second, the institution that holds the data must have procedures in place that govern who can access data encrypted using advanced privacy-enhancing technologies. Institutions should also determine which internal entity is responsible for governing access requests. These entities can include ethics review committees or data access committees [2].

Conclusion

Medical data sharing is essential for modern clinical practice and medical research. However, traditional privacy-preserving technologies based on data perturbation, along with centralized and decentralized data-sharing models, carry inherent privacy risks and may have high impact on data utility. These shortcomings mean that research and health care institutions combine these traditional privacy-preserving technologies with contractual mechanisms to govern data sharing and comply with data protection laws. These contractual mechanisms are context-dependent and require trusted environments between research and health care institutions. Although federated learning models can help alleviate these risks as only aggregate-level data are shared across institutions, there are still orthogonal risks to privacy from indirect reidentification of patients from partial results [66]. Furthermore, changes in case law (such as the already mentioned recent invalidation of the US-EU Privacy Shield [3]) can undermine data sharing with research partners outside the European Union. In this paper, we demonstrated how these privacy risks can be addressed through using multiparty homomorphic encryption, an efficient combination of homomorphic encryption and secure multiparty computation. In particular, we demonstrated how homomorphic encryption and secure multiparty computation can be used to compute accurate federated analytics without needing to transfer personal

data. Combining these technologies (multiparty homomorphic encryption) for medical data sharing can improve the performance overheads of privacy enhancing technology while reducing the risk of GDPR noncompliance. Furthermore, personal data do not leave the host institution where they are stored when processed using multiparty homomorphic encryption. Therefore, the lack of personal data transfer with multiparty homomorphic encryption will encourage increased data sharing and standardization between institutions. Data protection agencies, as well as health care and research institutions, should promote multiparty homomorphic encryption and other advanced privacy-enhancing technologies for their use to become widespread for clinical and research data sharing.

Acknowledgments

We are indebted to Dan Bogdanov, Brad Malin, Sylvain Métille, and Pierre Hutter for their invaluable feedback on earlier versions of this manuscript. This work was partially funded by the Personalized Health and Related Technologies Program (grant 2017-201; project: Data Protection and Personalized Health) supported by the Council of the Swiss Federal Institutes of Technology.

Conflicts of Interest

None declared.

References

1. Dove ES, Knoppers BM, Zawati MH. An ethics safe harbor for international genomics research? Genome Med 2013;5(11):99 [FREE Full text] [doi: 10.1186/gm503] [Medline: 24267880]
2. Phillips M, Molnár-Gábor F, Korbel J, Thorogood A, Joly Y, Chalmers D, et al. Genomics: data sharing needs an international code of conduct. Nature 2020 Feb 05;578(7798):31-33 [FREE Full text] [doi: 10.1038/d41586-020-00082-9]
3. Case C-311/2018 judgment of the court (grand chamber) of 16 July 2020. Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems. request for a preliminary ruling from the high court (Ireland). reference for a preliminary ruling — protection of individuals with regard to the processing of personal data — charter of fundamental rights of the European Union — articles 7, 8 and 47 — regulation (EU) 2016/679 — article 2(2) — scope — transfers of personal data to third countries for commercial purposes — article 45 — commission adequacy decision — article 46 — transfers subject to appropriate safeguards — article 58 — powers of the supervisory authorities — processing of the data transferred by the public authorities of a third country for national security purposes — assessment of the adequacy of the level of protection in the third country — decision 2010/87/EU — protective standard clauses on the transfer of personal data to third countries — suitable safeguards provided by the data controller — validity — implementing decision (EU) 2016/1250 — adequacy of the protection provided by the EU-US privacy shield — validity — complaint by a natural person whose data was transferred from the European Union to the United States). Eur-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62018CJ0311 [accessed 2021-02-10]
4. Ledford H, Van Noorden R. High-profile coronavirus retractions raise concerns about data oversight. Nature 2020 Jun 5;582(7811):160-160 [FREE Full text] [doi: 10.1038/d41586-020-01695-w] [Medline: 32504025]
5. Chen J. How the best-laid plans go awry: the (unsolved) issues of applicable law in the General Data Protection Regulation. Int Data Priv Law 2016 Dec 10;6(4):310-323 [FREE Full text] [doi: 10.1093/idpl/ipw020]
6. Ienca M, Scheibner J, Ferretti A, Gille F, Amann J, Sleigh J, et al. How the General Data Protection Regulation changes the rules for scientific research. ETH Zurich. 2019 Jul. URL: https://www.research-collection.ethz.ch/handle/20.500.11850/391622 [accessed 2021-02-12]
7. Murtagh MJ, Demir I, Jenkings KN, Wallace SE, Murtagh B, Boniol M, et al. Securing the data economy: translating privacy and enacting security in the development of DataSHIELD. Public Health Genomics 2012;15(5):243-253 [FREE Full text] [doi: 10.1159/000336673] [Medline: 22722688]
8. Bovenberg J, Peloquin D, Bierer B, Barnes M, Knoppers BM. How to fix the GDPR's frustration of global biomedical research. Science 2020 Oct 02;370(6512):40-42 [FREE Full text] [doi: 10.1126/science.abd2499] [Medline: 33004505]
9. Kierkegaard P. Electronic health record: wiring Europe's healthcare. Comput Law Secur Rev 2011 Sep 16;27(5):503-515 [FREE Full text] [doi: 10.1016/j.clsr.2011.07.013]
10. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. PLoS One 2011;6(12):e28071 [FREE Full text] [doi: 10.1371/journal.pone.0028071] [Medline: 22164229]
11. Lin Z, Owen AB, Altman RB. Genomic research and human subject privacy. Science 2004 Jul 09;305(5681):183 [FREE Full text] [doi: 10.1126/science.1095019] [Medline: 15247459]
12. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science 2013 Jan 18;339(6117):321-324 [FREE Full text] [doi: 10.1126/science.1229566] [Medline: 23329047]
13. Lippert C, Sabatini R, Maher MC, Kang EY, Lee S, Arikan O, et al. Identification of individuals by trait prediction using whole-genome sequencing data. Proc Natl Acad Sci U S A 2017 Sep 19;114(38):10166-10171 [FREE Full text] [doi: 10.1073/pnas.1711125114] [Medline: 28874526]

14. El Emam K, Buckeridge D, Tamblyn R, Neisa A, Jonker E, Verma A. The re-identification risk of Canadians from longitudinal demographics. BMC Med Inform Decis Mak 2011 Jun 22;11(1):1-12 [FREE Full text] [doi: 10.1186/1472-6947-11-46] [Medline: 21696636]
15. Loukides G, Denny JC, Malin B. The disclosure of diagnosis codes can breach research participants' privacy. J Am Med Inform Assoc 2010 May 01;17(3):322-327 [FREE Full text] [doi: 10.1136/jamia.2009.002725] [Medline: 20442151]
16. Atreya R, Smith J, McCoy A, Malin B, Miller R. Reducing patient re-identification risk for laboratory results within research datasets. J Am Med Inform Assoc Oxford Academic Jan 1 2013;20(1):95-101 [FREE Full text] [doi: 10.1136/amiajnl-2012-001026] [Medline: 22822040]
17. Austin L, Lie D. Safe Sharing Sites. N Y Univ Law Rev 2019;94(4):581-623 [FREE Full text]
18. Bot BM, Wilbanks JT, Mangravite LM. Assessing the consequences of decentralizing biomedical research. Big Data Soc 2019 Jun 11;6(1):1-6 [FREE Full text] [doi: 10.1177/2053951719853858]
19. Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet 2008 Aug 29;4(8):e1000167 [FREE Full text] [doi: 10.1371/journal.pgen.1000167] [Medline: 18769715]
20. Sankararaman S, Obozinski G, Jordan MI, Halperin E. Genomic privacy and limits of individual detection in a pool. Nat Genet 2009 Aug 23;41(9):965-967 [FREE Full text] [doi: 10.1038/ng.436] [Medline: 19701190]
21. Shringarpure SS, Bustamante CD. Privacy risks from genomic data-sharing beacons. Am J Hum Genet 2015 Nov 05;97(5):631-646 [FREE Full text] [doi: 10.1016/j.ajhg.2015.09.010] [Medline: 26522470]
22. Raisaro JL, Tramèr F, Ji Z, Bu D, Zhao Y, Carey K, et al. Addressing beacon re-identification attacks: quantification and mitigation of privacy risks. J Am Med Inform Assoc 2017 Jul 01;24(4):799-805 [FREE Full text] [doi: 10.1093/jamia/ocw167] [Medline: 28339683]
23. Liu Y, Wan Z, Xia W, Kantarcioglu M, Vorobeychik Y, Clayton EW, et al. Detecting the presence of an individual in phenotypic summary data. AMIA Annu Symp Proc 2018;2018:760-769 [FREE Full text] [Medline: 30815118]
24. Dwork C, McSherry F, Nissim K, Smith A. Calibrating noise to sensitivity in private data analysis. In: Halevi S, Rabin T, editors. Theory of Cryptography. Berlin, Heidelberg: Springer; 2006:265-284.
25. Wood A, Altman M, Bembenek A, Bun M, Gaboardi M, Honaker J, et al. Differential privacy: a primer for a non-technical audience. Vanderbilt J Entertain Technol Law 2019;21(1):209-276 [FREE Full text]
26. Jayaraman B, Evans D. Evaluating differentially private machine learning in practice. 2019 Presented at: 28th USENIX Security Symposium; 14-16 August 2019; Santa Clara p. 1895-1912 URL: https://www.usenix.org/conference/usenixsecurity19/presentation/jayaraman
27. Nasirigerdeh R, Torkzadehmahani R, Matschinske J, Frisch T, List M, Späth J, et al. sPLINK: a federated, privacy-preserving tool as a robust alternative to meta-analysis in genome-wide association studies. BioRxiv. Preprint posted online on June 6, 2020. [FREE Full text] [doi: 10.1101/2020.06.05.136382]
28. Kairouz P, McMahan H, Avent B, Bellet A, Bennis M, Bhagoji A, et al. Advances and open problems in federated learning. ArXiv. Preprint posted online on December 10, 2019. [FREE Full text] [doi: 10.1561/2200000083]
29. Gaye A, Marcon Y, Isaeva J, LaFlamme P, Turner A, Jones EM, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. Int J Epidemiol 2014 Dec;43(6):1929-1944 [FREE Full text] [doi: 10.1093/ije/dyu188] [Medline: 25261970]
30. Melie-Garcia L, Draganski B, Ashburner J, Kherif F. Multiple linear regression: bayesian inference for distributed and big data in the medical informatics platform of the human brain project. BioRxiv. Preprint posted online on January 5, 2018. [FREE Full text] [doi: 10.1101/242883]
31. Rigaki M, Garcia S. A survey of privacy attacks in machine learning. ArXiv. Preprint posted online on July 15, 2020. [FREE Full text]
32. Carlini N, Liu C, Erlingsson Ú, Kos J, Song D. The secret sharer: evaluating and testing unintended memorization in neural networks. 2019 Presented at: 28th USENIX Security Symposium; 14-16 August; Santa Clara URL: https://www.usenix.org/conference/usenixsecurity19/presentation/carlini
33. Hitaj B, Ateniese G, Perez-Cruz F. Deep models under the GAN: information leakage from collaborative deep learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017 Presented at: ACM Conference on Computer and Communications Security (CCS) 2017; October 30-November 3; Dallas p. 603-618. [doi: 10.1145/3133956.3134012]
34. Nasr M, Shokri R, Houmansadr A. Comprehensive privacy analysis of deep learning: passive and active white-box inference attacks against centralized and federated learning. 2019 Presented at: 2019 IEEE Symposium on Security and Privacy; May 19-23; San Francisco p. 739-753 URL: https://ieeexplore.ieee.org/abstract/document/8835245 [doi: 10.1109/sp.2019.00065]
35. Melis L, Song C, De CE, Shmatikov V. Exploiting unintended feature leakage in collaborative learning. 2019 Presented at: 2019 IEEE Symposium on Security and Privacy; May 19-23; San Francisco p. 691-706 URL: https://ieeexplore.ieee.org/abstract/document/8835269 [doi: 10.1109/sp.2019.00029]
36. Wallace SE, Gaye A, Shoush O, Burton PR. Protecting personal data in epidemiological research: DataSHIELD and UK law. Public Health Genomics 2014;17(3):149-157 [FREE Full text] [doi: 10.1159/000360255] [Medline: 24685519]

37. Gentry C. Computing arbitrary functions of encrypted data. Commun ACM 2010 Mar;53(3):97-105 [FREE Full text] [doi: 10.1145/1666420.1666444]
38. Yao A. Protocols for secure computations. 1982 Presented at: 23rd Annual Symposium on Foundations of Computer Science; Nov 3-5; Chicago p. 160-164 URL: https://ieeexplore.ieee.org/abstract/document/4568388 [doi: 10.1109/sfcs.1982.38]
39. Bogdanov D, Laur S, Willemson J. Sharemind: a framework for fast privacy-preserving computations. In: Jajodia S, Lopez J, editors. Computer Security European Symposium on Research in Computer Security. Berlin: Springer; 2008:192-206.
40. Damgård I, Pastro V, Smart N, Zakarias S. Multiparty computation from somewhat homomorphic encryption. In: Safavi-Naini R, Canetti R, editors. Advances in Cryptology CRYPTO Lecture Notes in Computer Science vol 7412. Berlin: Springer; 2012:643-662.
41. Damgård I, Keller M, Larraia E, Pastro V, Scholl P, Smart N. Practical covertly secure MPC for dishonest majority - or: breaking the SPDZ limits. In: Crampton J, Jajodia S, Mayes K, editors. Computer Security European Symposium on Research in Computer Security. Berlin: Springer; 2013:1-18.
42. Keller M, Pastro V, Rotaru D. Overdrive: making SPDZ great again. In: Nielsen JB, Rijmen V, editors. Advances in Cryptology EUROCRYPT Lecture Notes in Computer Science vol 10822. Cham: Springer International Publishing; 2018:158-189.
43. Mouchet C, Troncoso-Pastoriza J, Bossuat JP, Hubaux JP. Multiparty homomorphic encryption: from theory to practice. Cryptology ePrint Archive. 2020 Dec 22. URL: https://eprint.iacr.org/2020/304 [accessed 2021-02-12]
44. Cramer R, Damgård I, Nielsen J. Multiparty computation from threshold homomorphic encryption. In: Pfitzmann B, editor. Advances in Cryptology EUROCRYPT 2001 Lecture Notes in Computer Science vol 2045. Berlin: Springer; 2001:280-300.
45. Asharov G, Jain A, López-Alt A, Tromer E, Vaikuntanathan V, Wichs D. Multiparty computation with low communication, computation and interaction via threshold FHE. In: Pointcheval D, Johansson T, editors. Advances in Cryptology EUROCRYPT 2012 Lecture Notes in Computer Science vol 7237. Berlin: Springer; 2012:483-501.
46. Boneh D, Gennaro R, Goldfeder S, Jain A, Kim S, Rasmussen P, et al. Threshold cryptosystems from threshold fully homomorphic Eencryption. In: Shacham H, Boldyreva A, editors. Advances in Cryptology CRYPTO Lecture Notes in Computer Science vol 10991. Cham: Springer International Publishing; 2018:565-596.
47. Zheng W, Popa R, Gonzalez J, Stoica I. Helen: maliciously Ssecure coopetitive learning for linear models. 2019 Presented at: 2019 IEEE Symposium on Security and Privacy; May 19-23; San Francisco p. 724-738 URL: https://ieeexplore.ieee.org/abstract/document/8835215 [doi: 10.1109/sp.2019.00045]
48. Raisaro JL, Troncoso-Pastoriza JR, Misbach M, Sousa JS, Pradervand S, Missiaglia E, et al. MedCo: enabling secure and privacy-preserving exploration of distributed clinical and genomic data. IEEE/ACM Trans Comput Biol and Bioinf 2019 Jul 1;16(4):1328-1341. [doi: 10.1109/tcbb.2018.2854776]
49. Sav S, Pyrgelis A, Troncoso-Pastoriza J, Froelicher D, Bossuat JP, Sousa J, et al. POSEIDON: privacy-preserving federated neural network learning. arXiv. Preprint posted online on January 8, 2021. [FREE Full text]
50. Spindler G, Schmechel P. Personal data and encryption in the European General Data Protection Regulation. J Intellect Prop Inf Technol Electron Commer Law 2016;7(2):163 [FREE Full text]
51. Finck M, Pallas F. They who must not be identified - distinguishing personal from non-personal data under the GDPR. Int Data Priv Law 2020 Feb 1;10(1):11-36 [FREE Full text] [doi: 10.1093/idpl/ipz026]
52. Guidance on anonymisation and pseudonymisation. Coimisiún Chosaint Sonraí Data Protection Commission. 2019. URL: https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf [accessed 2021-02-12]
53. Orientierungshilfe zum Gesundheitsdatenschutz. Bundesministerium für Wirtschaft und Energie. 2018 Nov. URL: https://www.bmwi.de/Redaktion/DE/Downloads/M-O/orientierungshilfe-gesundheitsdatenschutz.pdf?__blob=publicationFile&v=16 [accessed 2021-02-12]
54. Article 29 Data Protection Working Party. Opinion 04/2007 on the concept of personal data. European Commission. 2007. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf [accessed 2021-02-18]
55. Article 29 Data Protection Working Party. Opinion 05/2014 on anonymisation techniques. European Commission. 2014 Apr. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf [accessed 2021-02-18]
56. Esayas S. The role of anonymisation and pseudonymisation under the EU data privacy rules: beyond the 'all or nothing' approach. Eur J Law Technol Vol 2015;6(2):1-28 [FREE Full text]
57. Nissim K, Bembenek A, Wood A, Bun M, Gaboardi M, Gasser U, et al. Bridging the gap between computer science and legal approaches to privacy. Harv J Law Technol 2018;31(2):2017-2780 [FREE Full text]
58. Case C-582/14 judgement of the court (second chamber) of 19 October 2016 (request for a preliminary ruling from the Bundesgerichtshof – Germany) – Patrick Breyer v Bundesrepublik Deutschland. Eur-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62014CA0582&qid=1614053861638 [accessed 2021-02-10]
59. Mondschein C, Monda C. The EU's General Data Protection Regulation (GDPR) in a research context. In: Kubben P, Dumontier M, Dekker A, editors. Fundamentals of Clinical Data Science. Cham: Springer; 2019:55-71.

60. Case C-210/16 judgment of the court (grand chamber) of 5 June 2018 (request for a preliminary ruling from the Bundesverwaltungsgericht — Germany) — Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH. Eur-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62016CA0210&qid=1613633585608 [accessed 2021-02-18]
61. Case C-25/17 judgment of the court (grand chamber) of 10 July 2018 (request for a preliminary ruling from the korkein hallinto-oikeus — Finland) — proceedings brought by Tietosuojavaltuutettu. Eur-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62017CA0025&qid=1613633621924 [accessed 2021-02-18]
62. Case C-40/17 judgement of the court (grand chamber) of 29 July 2019 (request for a preliminary ruling from the Oberlandesgericht Düsseldorf — Germany) – Fashion ID GmbH & Co. KG v Verbraucherzentrale NRW eV. Eur-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62017CA0040&qid=1613634205133 [accessed 2021-02-18]
63. Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (General Data Protection Regulation). Eur-Lex. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679&qid=1613633727166 [accessed 2021-02-18]
64. 4A_365/2017. Tribunal fédéral. URL: https://www.bger.ch/ext/eurospider/live/fr/php/aza/http/index.php?lang=fr&type=show_document&highlight_docid=aza://26-02-2018-4A_365-2017 [accessed 2021-02-18]
65. Staiger DN. Swiss data protection law. In: Moura Vicente D, de Vasconcelos Casimiro S, editors. Data Protection in the Internet Ius Comparatum - Global Studies in Comparative Law vol 38. Cham: Springer International Publishing; 2020:397-408.
66. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nat Mach Intell 2020 Jun 8;2(6):305-311 [FREE Full text] [doi: 10.1038/s42256-020-0186-1]
67. Jagadeesh KA, Wu DJ, Birgmeier JA, Boneh D, Bejerano G. Deriving genomic diagnoses without revealing patient genomes. Science 2017 Aug 18;357(6352):692-695 [FREE Full text] [doi: 10.1126/science.aam9710] [Medline: 28818945]
68. Munn L, Hristova T, Magee L. Clouded data: privacy and the promise of encryption. Big Data Soc 2019 Jun 24;6(1):1-16 [FREE Full text] [doi: 10.1177/2053951719848781]
69. Grishin D, Raisaro J, Troncoso-Pastoriza J, Obbad K, Quinn K, Misbach M, et al. Citizen-centered, auditable, and privacy-preserving population genomics. bioRxiv. Preprint posted online on October 10, 2019. [FREE Full text] [doi: 10.1101/799999]
70. Albrecht M, Chase M, Chen H, Ding J, Goldwasser S, Gorbunov S, et al. Homomorphic encryption standard internet report no. 939. Cryptology ePrint Archive. 2019 Aug 17. URL: https://eprint.iacr.org/2019/939 [accessed 2021-02-12]
71. A new guide regarding security of personal data. Commission Nationale de l'Informatique et des Libertés. 2018 Apr 4. URL: https://www.cnil.fr/en/new-guide-regarding-security-personal-data [accessed 2021-02-12]
72. Encryption and privacy III: homomorphic encryption. Agencia Española de Protección de Datos. 2020. URL: https://www.aepd.es/en/prensa-y-comunicacion/blog/encryption-privacy-iii-homomorphic-encryption [accessed 2021-02-12]
73. Laine K. Homomorphic encryption. In: Jiang X, Tang H, editors. Responsible Genomic Data Sharing Challenges and Approaches. Cambridge: Academic Press; 2019:97-122.
74. Vinterbo SA, Sarwate AD, Boxwala AA. Protecting count queries in study design. J Am Med Inform Assoc 2012;19(5):750-757 [FREE Full text] [doi: 10.1136/amiajnl-2011-000459] [Medline: 22511018]

Abbreviations

CJEU: Court of Justice of the European Union
COVID-19: coronavirus disease 2019
GDPR: (European Union) General Data Protection Regulation
HIV: human immunodeficiency virus

Edited by G Eysenbach; submitted 19.10.20; peer-reviewed by B Knoppers, S Palmdorf, R Hendricks-Sturrup; comments to author 07.12.20; revised version received 06.01.21; accepted 16.01.21; published 25.02.21.

Please cite as:
Scheibner J, Raisaro JL, Troncoso-Pastoriza JR, Ienca M, Fellay J, Vayena E, Hubaux JP
Revolutionizing Medical Data Sharing Using Advanced Privacy-Enhancing Technologies: Technical, Legal, and Ethical Synthesis
J Med Internet Res 2021;23(2):e25120
URL: https://www.jmir.org/2021/2/e25120
doi:10.2196/25120
PMID:33629963

Review

Associations Between Digital Health Intervention Engagement, Physical Activity, and Sedentary Behavior: Systematic Review and Meta-analysis

Matthew Mclaughlin[1,2,3,4], BSc; Tessa Delaney[1,2,3,4], BND; Alix Hall[1,2,3,4], PhD; Judith Byaruhanga[1,2,3,4], MPH; Paul Mackie[5,6], MSc; Alice Grady[1,2,3,4], PhD; Kathryn Reilly[1,2,3,4], PhD; Elizabeth Campbell[1,2,3,4], PhD; Rachel Sutherland[1,2,3,4], PhD; John Wiggers[1,2,3,4], PhD; Luke Wolfenden[1,2,3,4], PhD

[1]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

[2]Hunter New England Population Health, Wallsend, Australia

[3]Hunter Medical Research Institute, New Lambton Heights, Australia

[4]Priority Research Centre for Heath Behaviour, University of Newcastle, Callaghan, Australia

[5]School of Health Sciences and Priority Research Centre for Stroke and Brain Injury, University of Newcastle, Callaghan, Australia

[6]Centre for Research Excellence in Stroke Recovery and Rehabilitation, Florey Institute of Neuroscience, Melbourne, Australia

Corresponding Author:
Matthew Mclaughlin, BSc
School of Medicine and Public Health
University of Newcastle
University Drive
Callaghan, 2308
Australia
Phone: 61 02 4924 6477
Email: Matthew.Mclaughlin1@health.nsw.gov.au

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/3/e29094

Abstract

Background: The effectiveness of digital health interventions is commonly assumed to be related to the level of user engagement with the digital health intervention, including measures of both digital health intervention use and users' subjective experience. However, little is known about the relationships between the measures of digital health intervention engagement and physical activity or sedentary behavior.

Objective: This study aims to describe the direction and strength of the association between engagement with digital health interventions and physical activity or sedentary behavior in adults and explore whether the direction of association of digital health intervention engagement with physical activity or sedentary behavior varies with the type of engagement with the digital health intervention (ie, subjective experience, activities completed, time, and logins).

Methods: Four databases were searched from inception to December 2019. Grey literature and reference lists of key systematic reviews and journals were also searched. Studies were eligible for inclusion if they examined a quantitative association between a measure of engagement with a digital health intervention targeting physical activity and a measure of physical activity or sedentary behavior in adults (aged ≥18 years). Studies that purposely sampled or recruited individuals on the basis of pre-existing health-related conditions were excluded. In addition, studies were excluded if the individual engaging with the digital health intervention was not the target of the physical activity intervention, the study had a non–digital health intervention component, or the digital health interventions targeted multiple health behaviors. A random effects meta-analysis and direction of association vote counting (for studies not included in meta-analysis) were used to address objective 1. Objective 2 used vote counting on the direction of the association.

Results: Overall, 10,653 unique citations were identified and 375 full texts were reviewed. Of these, 19 studies (26 associations) were included in the review, with no studies reporting a measure of sedentary behavior. A meta-analysis of 11 studies indicated

a small statistically significant positive association between digital health engagement (based on all usage measures) and physical activity (0.08, 95% CI 0.01-0.14, SD 0.11). Heterogeneity was high, with 77% of the variation in the point estimates explained by the between-study heterogeneity. Vote counting indicated that the relationship between physical activity and digital health intervention engagement was consistently positive for three measures: subjective experience measures (2 of 3 associations), activities completed (5 of 8 associations), and logins (6 of 10 associations). However, the direction of associations between physical activity and time-based measures of usage (time spent using the intervention) were mixed (2 of 5 associations supported the hypothesis, 2 were inconclusive, and 1 rejected the hypothesis).

Conclusions: The findings indicate a weak but consistent positive association between engagement with a physical activity digital health intervention and physical activity outcomes. No studies have targeted sedentary behavior outcomes. The findings were consistent across most constructs of engagement; however, the associations were weak.

(J Med Internet Res 2021;23(2):e23180) doi:10.2196/23180

KEYWORDS

engagement; adherence; digital health intervention; digital behavior change intervention; physical activity; sedentary behavior; mobile phone

Introduction

Physical activity of any intensity reduces the risk of death and noncommunicable diseases [1]. Sedentary behavior is highly prevalent, displaces time to be physically active, and is associated with noncommunicable diseases and premature death [2,3]. As such, efforts to increase physical activity and concurrently decrease sedentary behavior have been identified internationally as public health priorities [4].

Digital health interventions (DHIs) have the potential to address physical inactivity, as they are accessible by large proportions of the population and can be delivered with high effectiveness at a low cost [5,6]. The World Health Organization defines *digital health* as the use of digital, mobile, and wireless technologies to support the achievement of health objectives and is inclusive of both mobile health (mHealth) and eHealth [7], including mobile phones, portable computer tablets (eg, iPads), web-based interventions, smartphone apps, and wearable devices [8]. An attractive feature of DHIs is their capacity to be scaled for large populations while concurrently being tailored to specific target groups [7,9]. There has been a considerable investment in the development and research on DHIs to improve physical activity, with a rise of 26% per year in journal article publications since 2000 [10]. Furthermore, the use of DHIs to promote and support participation in physical activity has been recommended in the Global Action Plan on Physical Activity 2018-2030 [4,11].

Despite the significant growth in DHIs, there is a limited understanding of the extent to which DHIs impact physical activity outcomes. Overall, systematic reviews indicate that DHIs targeting adult populations may have a modest effect in improving physical activity when delivered web based [12-16] and through smartphone applications [12,13,17-19]. A recent meta-analysis found that per week, web-based interventions increased moderate-to-vigorous intensity physical activity by 13.4 minutes (95% CI 12.96-13.89) and steps by 2185 (95% CI 1765-2605) [15]. Similarly, another meta-analysis of web-based interventions targeting physical activity found that physical activity significantly improved in the short term (Cohen *d*=0.14). The review suggested that the small effect size may be attributed to a lack of engagement with web-based interventions [14]. Other meta-analyses of mobile phone app interventions have reported effects that favored the intervention but were not statistically significant [17,19], with the suggestion that a lack of engagement may explain the lack of evidence to support effectiveness.

Participant exposure needs to be sufficient for any DHI to have an effect [20]. Engagement has been defined as the (1) extent of DHI *usage* such as the frequency, duration, amount, and depth of the accessed DHI and (2) *subjective experience* characterized by attention, interest, and affect [20]. To our knowledge, only one systematic review has explored the association between objective levels of engagement with DHIs (usage measures) and physical activity or sedentary behavior [8]. The review by Donkin et al [8] explored the association between the level of engagement with DHIs (web-based interventions) targeting adults and a range of health outcomes [8]. Included studies predominately reported the measures of psychological health, dietary intake, weight management, and smoking. The results were reported narratively because of the diverse measures of engagement and health outcomes. Only 3 of the 33 studies included measures of physical activity (n=5 associations) [21-24], with no studies reporting sedentary behavior outcomes. The engagement measures explored by the 3 studies included in the Donkin review focused on logins (n=3), activities completed (n=1), and website exposure (n=1). Of the 3 physical activity studies, Marcus et al [22] showed that a higher number of logins were correlated with an increase in the physical activity from baseline to 12 months. Similarly, McKay et al [21] found that those who logged into the program on 3 or more occasions had greater increases in physical activity than those with fewer logins. McKay et al [21] additionally found website exposure (usage) to be associated with higher increases in physical activity. In contrast, Carr et al [23,24] showed that neither the number of logins nor the number of activities completed were associated with physical activity at 8 months.

Understanding the relationship between engagement and health outcomes is important because it provides an opportunity to optimize the impact of interventions [20,25]. Engagement is hypothesized to influence the relationship between a DHI and the mechanisms of action of the DHI (eg, skills, attitudes, beliefs, knowledge), which then leads to the target behavior (eg,

physical activity) [20]. Even smaller improvements in the effectiveness of DHIs are important, given the potential reach of these interventions [6,10]. The previous review by Donkin et al [8] used a definition of engagement that focused on the usage and user-directed web-based interventions, which excluded smartphone apps and group-based DHIs [8]. Since 2010, when their search was conducted, there has been a large increase in the mobile- and app-based research applied to physical activity; hence, there is an increased opportunity to garner further understanding of this relationship [8,10]. A broader definition of engagement that encompasses subjective experience has also been developed [20]. To our knowledge, no review has explored the relationship between subjective experience with DHIs and physical activity or sedentary behavior [20]. Therefore, a more contemporary review of the evidence is warranted.

Objective

In this context, we aim to (1) describe the direction and strength of the association between engagement with DHIs and physical activity and/or sedentary behavior in adults and (2) explore whether the direction of association between DHI engagement and physical activity or sedentary behavior varies by the type of engagement (ie, subjective experience, activities completed, time, and logins).

Methods

Design

This review was prospectively registered with the International Prospective Register for Systematic Reviews (PROSPERO; CRD42018110657) and is reported in accordance with the Joanna Briggs Institute guidance for conducting systematic reviews of association [26].

Search Strategy

Searches for peer-reviewed literature were undertaken with the assistance of a research librarian in 4 electronic databases: Embase, MEDLINE, PsycINFO, and Scopus (Multimedia Appendix 1). We searched for records from the database inception to December 2019. Searches were restricted to English. This review was conducted alongside another review that aimed to describe the association between DHI engagement and dietary intake (PROSPERO CRD42018112189 [27]). Therefore, *dietary intake* search terms were also included in the search strategy. We used the modified versions of published search filters for physical activity and sedentary behavior [28], engagement [20], and DHIs [20,29,30].

Additional Search Methods

We conducted hand searches of all the publications from January 2016 to December 2019 in the following journals: *Journal of Medical Internet Research*, *JMIR mHealth and uHealth*, *JMIR Medical Informatics*, and *JMIR Public Health and Surveillance*. We conducted gray literature searches in "Google.com/ncr" search engine and used the search terms *Physical Activity* or *Sedentary Behavior* and *Engagement* and *Digital Health Intervention* and screened the first 200 hits for relevance. We screened the reference lists of key systematic reviews of DHI engagement [8,20]. We also contacted the authors of included studies for other potentially relevant studies.

Inclusion Criteria

Types of Studies

We included study designs that examined a quantitative association between a measure of engagement with a DHI and a measure of physical activity and/or sedentary behavior. DHIs were defined as the use of digital, mobile, and wireless technologies to support the achievement of health objectives [7], inclusive of both mHealth and eHealth. We adopted Perski et al's [20] definition of engagement, defined as both the extent of the usage of the DHI (amount, frequency, duration, and depth; eg, activities completed, time, and logins) and the subjective experience (characterized by attention, interest, and affect). DHIs included but were not limited to mobile phones, portable computer tablets (eg, iPads), web-based interventions, and smartphone apps. We included DHIs involving synchronous communication as part of the program (eg, web-based chat, teleconferencing). We applied no restrictions on the length of the follow-up period or the country of origin of the studies. We included studies that recruited participants in the real-world settings (ie, ecological studies) as well as nonecological studies (ie, those conducted under controlled research conditions, where repeated contacts with research staff, comprehensive assessments, and recruitment to the study occurs before the individual accessing the DHI) [31]. All the quantitative study designs were also included.

Population

We included any study undertaken with adult users (aged ≥18 years) of a DHI targeting physical activity or sedentary behavior. The studies of participants who had access to a DHI and the opportunity to engage with the DHI were eligible.

Exposure

We included studies reporting any measure of engagement with a DHI, defined as the extent of usage (eg, activities completed, time, and logins) or the subjective experience of users (eg, measures of attention, interest, and affect, including but not limited to enjoyment, satisfaction, user experience, and usability) [20]. Engagement can be collected by the DHI (eg, analytics), observation, surveys of DHI users, or other quantitative methods. We excluded the qualitative measures of engagement (eg, focus groups).

Outcome

We included studies reporting any measure of physical activity or sedentary behavior, including but not limited to self-report (eg, minutes of moderate-to-vigorous physical activity, minutes of walking, self-reported steps, distance traveled) and measured by a device (eg, steps from pedometer, mobile phone data, accelerometers). These could be reported in specific settings (eg, while at work), periods of the day (eg, mornings), or as the whole day. We included both cross-sectional measures of physical activity (ie, one time point) and those studies with multiple time points calculating changes in the physical activity over time (ie, cohort studies).

Exclusion Criteria

We excluded the following studies:

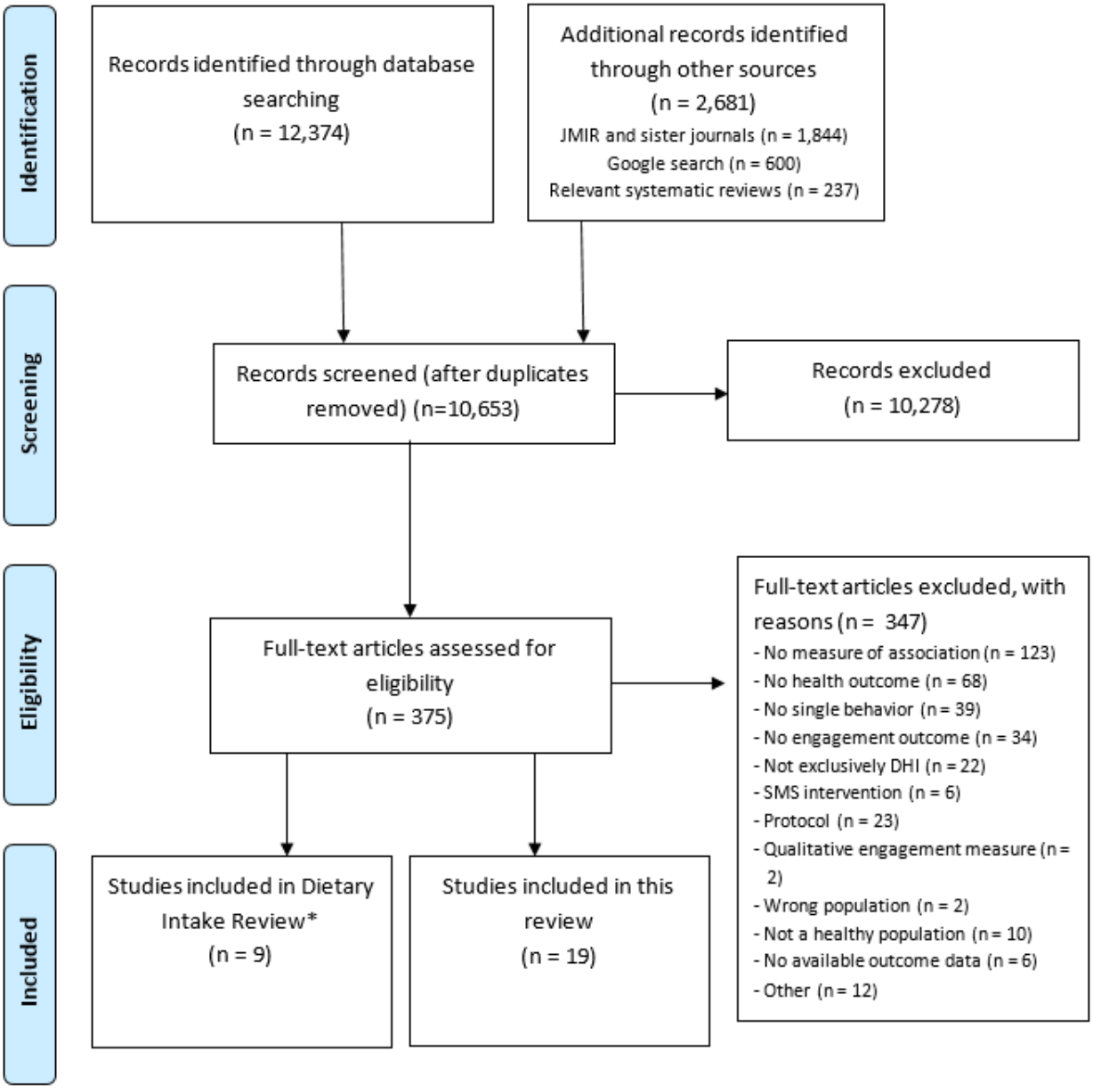
- Case studies, letters to the editor, and qualitative studies.
- Studies that targeted children and adolescents (<18 years).
- Studies that purposely sampled or recruited individuals based on pre-existing health-related conditions, including chronic health conditions such as chronic pain, a diagnosis of chronic disease, communicable disease, or mental illness, given our interest in generalizing the findings to general community samples.
- Studies in which the individuals engaging in the DHI were not the target of the physical activity or sedentary behavior intervention (eg, doctors engaging with a physical activity app for their patients).
- Studies that included a non-DHI component within an intervention (eg, a face-to-face component and digital components). This step was taken to ensure that the measures of engagement reflected only the digital component and not the intervention more generally.
- Interventions not functioning at computer- or internet-based capacity (eg, SMS, CD-Rom, and computer-based interventions) to focus on more contemporary DHIs.
- Those that targeted multiple health behaviors for the prevention of chronic disease (eg, sleep and physical activity or diet and physical activity) to reduce heterogeneity between health behaviors.
- Studies where the full text was not available.

Data Collection and Analyses

Selection of Studies

After removing the duplicates, the authors (MM, TD, and JB) single-screened titles and abstracts for potentially eligible studies using Covidence. At title and abstract screening, we included studies in the full text review when the abstract reported both a physical activity and/or sedentary behavior outcome as well as a DHI engagement outcome (including meeting other inclusion and exclusion criteria). Therefore, studies that did not report a measure of association between physical activity and/or sedentary behavior and DHI engagement were still included for a full text review. This was done to ensure that the studies were not excluded in error. This screening process was implemented after an initial pilot screening of 100 full texts by MM, who found that none of the abstracts that reported only a health outcome or only an engagement outcome were incorrectly screened out. Following the title and abstract review, we obtained full texts of all potentially relevant or unclear articles, and authors (MM, TD, AG, and KR) independently reviewed these against our inclusion criteria. Reasons for exclusion were recorded in a *characteristics of excluded studies* table. The review authors were not blinded to author or journal information. The number of articles identified, screened, eligible, and included were recorded according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [32] (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. *This review was conducted alongside another review aiming to describe the association between DHI engagement and dietary intake (PROSPERO CRD42018112189). Therefore, 'dietary intake' search terms were also included in the search strategy, but removed for reporting elsewhere.



Data Extraction and Management

Pairs of review authors (MM and TD) independently extracted data using a data extraction form adapted from the Cochrane Public Health Group Methods Manual and used previously by the research team [33]. Given the complexity of the review, all the extracted data were reviewed by an experienced statistician and methods expert, and any disagreements were resolved by the third author (AH). In case of missing study data, we attempted to contact the corresponding authors to obtain the required information. The following data were extracted:

- Study characteristics including authors' name, year of publication, overall study design, intervention target (ie, intended intervention recipient), recruitment method (ecological or nonecological), and sample size.
- Characteristics of the intervention, including type of DHI (ie, web-based, smartphone app, Exergame, and Facebook group), description of DHI components, and length of exposure to DHI.
- Outcomes including both a description of the physical activity or sedentary behavior measure and the engagement measure; the analysis method used to examine associations including adjustments for confounding (for quality assessment); magnitude of the association (ie, odds ratio [OR] or regression coefficient or estimate along with a measure of variability [if available], 95% CIs or standard deviation or standard error); statistical significance of the association; and further information to allow quality assessment.

Critical Appraisal

Pairs of review authors (MM, PM, and RS) assessed the methodological quality of studies independently using the Newcastle Ottawa Scale for cohort studies [34] (n=13 studies) and cross-sectional studies [35] (n=6 studies). We defined cross-sectional studies as the ones using a single time point of data for measuring the physical activity or sedentary behavior (eg, follow-up), whereas cohort studies as those that used multiple time points of data and calculated changes over time (eg, change from baseline to follow-up). All the studies were assessed based on their highest quality measure of association (ie, final follow-up time point, adjusted, device-measured physical activity or sedentary behavior, and objectively assessed engagement were preferred to mid-point follow-up, unadjusted, self-reported physical activity or sedentary behavior, and self-reported engagement).

The Newcastle Ottawa Scales utilizes a star system to assess the methodological quality of studies. The cohort tool assigns a maximum of 9 points for quality assessment in 3 domains: (1) selection of study groups (up to 4 points), (2) comparability of these groups (up to 2 points), and (3) assessment of outcomes (up to 3 points). The cross-sectional tool assigns a maximum of 10 points across the same 3 domains: (1) selection of study groups (up to 5 points), (2) comparability of these groups (up to 2 points), and (3) assessment of outcomes (up to 3 points; Multimedia Appendices 2 and 3 for scoring systems, adopted from Wells et al [34] and Modesti et al [35]).

Within the cohort tool, for the item "was follow up long enough for outcomes to occur," studies were awarded a point if they had a minimum follow-up period of 3 months. This period was chosen based on the current evidence on the length of DHI engagement, where those that are designed to be used for 3 months or longer tend to be more effective than those designed for shorter durations [14,15]. Within the cross-sectional tool, for the item "the study controls for the most important factor (select one)," we selected age as the factor to control for, as it is unanimously possible to control for this across studies and is an important contextual factor influencing engagement [20]. Disagreements between assessments were resolved by discussion between the pairs of review authors (MM, RS, and PM) and, where required by consulting the third review author.

Data Synthesis and Analysis

Data were synthesized according to the review objectives.

Objective 1

> *Describe the direction and strength of the association between engagement with DHIs and physical activity and/or sedentary behavior in adults.*

We planned two separate meta-analyses: first for subjective experience and second for usage (including activities completed, time, and logins). However, we did not conduct a meta-analysis for subjective experience because of the considerable methodological heterogeneity between studies and the small number of studies reporting this outcome (n=3) [36]. Therefore, we focused on the direction and strength of the association between usage and physical activity or sedentary behavior.

A variety of different methods of association were used across the included studies. Consequently, to allow for meta-analysis, we were required to transform a number of estimates into one consistent effect index. A standardized regression coefficient was chosen as the effect index, which is a previously proposed method [37]. A list of the main transformations used for this analysis are detailed in Multimedia Appendix 4 and are predominantly based on the research works of Borenstein et al [36] and Nieminen et al [37].

We used the Dersimonian and Laird random effects method of meta-analysis to calculate a pooled standardized effect assessing the strength and direction of associations. Statistical analyses were performed using R [38]. Many studies have reported more than one association. For meta-analyses, we used the following hierarchical selection criteria to select a single association from each study for inclusion in the pooled synthesis:

- Use measures were given preference in the following order: activities completed, time on site then logins. This attribute reflects the level of participant involvement required for each measure of engagement, with a greater level of engagement given priority.
- Measures of total physical activity were preferred to specific physical activity or sedentary behavior types (eg, measures of physical activity of moderate-to-vigorous intensity were preferred to distance walked).
- Device-measured physical activity was preferred over self-report.
- Whole-day measures of physical activity or sedentary behavior were preferred over specific time segments (eg, whole-day physical activity was preferred to workday physical activity).
- Global scores of subjective experience engagement were preferred over the individual constructs of subjective experience (eg, engagement questionnaire score preferred to game-flow score only).
- For DHIs with team or group engagement opportunities (eg, group challenges in a step-counting website), individual measures of engagement (eg, individual logins to the DHI) were preferred to group opportunities to engage (eg, total logins from a team of DHI users), as this more accurately reflects an individual's DHI engagement.
- Associations derived from fully adjusted models were preferred over unadjusted or partly adjusted models.

When studies did not provide sufficient data required for meta-analysis (ie, information to calculate an effect estimate and measure of variability of the effect estimate), the corresponding authors were contacted via email on up to 3 occasions and asked to provide information. When data were not available or not provided, we excluded the study from the meta-analysis. To provide supplementary data for this objective, we used vote-counting synthesis methods to describe the direction of association across these studies. We used the direction of association rather than statistical significance in accordance with the recent SWIM (Synthesis Without Meta-Analysis) guidelines [39]. We focused on a single measure of association from each study, which was selected based on the same hierarchical criteria for the selection of an association as for the meta-analysis. For vote counting, each study was

summarized as either "+," "−" or "0." "+" was assigned to the studies in which the point estimate and CI supported the hypothesis that higher engagement is associated with a higher physical activity or reduced sedentary behavior. "0" was assigned to the studies in which the point estimate and CI had inconclusive findings. "−" was assigned to the studies in which the association point estimate and CI rejected the hypothesis. Several studies did not report any results, including an estimate or CI, but rather just stated whether the association was statistically significant in the hypothesized direction. In these instances, to ensure that such studies were included, we assigned either "+" or "−" to the studies that reported *significant* association findings depending on the stated direction of the association (if provided) or "0" to the studies that reported *nonsignificant findings*.

Objective 2

> *To explore whether the direction of association between engagement and physical activity or sedentary behavior varies according to the type of engagement (ie, subjective experience, activities completed, time, and logins).*

To explore whether the direction or strength of associations varied between the different types of engagement, we classified each association across all the studies as either subjective experience, activities completed, time (time spent using the intervention; eg, session duration), or logins. Studies could contribute more than one association but only one association for each of the four types of engagement.

When studies had more than one association for a given type of engagement, we gave preference to the measures of association from adjusted associations and excluded the measures of association from the self-report measures of physical activity or sedentary behavior where device measures were available for the same engagement variable. Finally, for studies reporting associations at multiple time points, we included only data from the final time point.

We used vote-counting methods to explore the direction of the association between each type of engagement and physical activity or sedentary behavior outcomes. Each association was summarized as either "+," "−," or "0," following the same procedures described above, with "0" being assigned to studies with associations reporting mixed findings.

Results

Search Results

The searches resulted in 13,192 potentially relevant abstracts. After removing the duplicates, 10,653 unique citations were retained for review. After the title and abstract screening, 375 full texts were identified and screened. Overall, 19 studies were included in our review (Figure 1).

Study Characteristics

Detailed characteristics of each study (n=19) [22,24,40-56], and each measure of association, are provided in Multimedia Appendix 5. All 19 studies (n=7776 participants) were of physical activity measures, with no study using measures of sedentary behavior. Of the 19 studies, 12 were web-based interventions [22,24,41,44,45,47-50,52,54,55], 5 were app-based [43,46,51,53,56], and the remaining were Facebook-based (n=1) [42] and exergames (n=1) [40]. Cohort designs were used in 13 studies [22,24,42-44,47,48,50-52,54-56]. The remaining 6 studies used cross-sectional designs [40,41,45,46,49,53]. Across both cohort and cross-sectional studies, 11 studies included an analysis of the intervention arm of a randomized controlled trial [22,24,42-45,48,50,52,54,55].

The majority of studies included the usage measures of engagement (ie, activities completed, time, or logins; n=18), whereas 3 studies included subjective experience measures of engagement [40,46,48]. Participants across all the studies were predominately female (71%). Most studies used nonecological recruitment methods (n=11), with the remaining 8 studies using a mixture of ecological and nonecological recruitment methods [17,44,46,49,51,53,55,56]. The sample sizes across all the studies ranged from 7 to 3555 (mean 389; SD 760.6).

Critical Appraisal (Quality Assessment)

Of the 19 studies, almost half were assessed to be of *poor* quality (n=9). Two studies were considered to be of fair quality, and the remaining studies (n=8) were considered to be of good quality. Quality assessment results for cohort studies are summarized in Table 1, and quality assessment results for cross-sectional studies are summarized in Table 2.

Table 1. Quality assessment (Newcastle-Ottawa Quality Assessment Scale criteria for cohort studies).

Study	Selection				Comparability	Outcome			Selection[a]
	Representativeness of the exposed cohort	Selection of the nonexposed cohort	Ascertainment of exposure	Outcome not present at the start of study	Cohort statistical analysis	Assessment of the outcome	Was follow-up long enough for outcomes to occur	Adequacy of follow-up cohorts	
Carr et al [24]	★	★	★	★	0	0	★	0	Poor
Edney et al [42]	★	★	★	0	0	0	0	★	Poor
Edney et al [43]	★	★	★	★	★	★	★	★	Good
Ferney et al [44]	★	★	★	0	★	0	★	★	Good
Kwan et al [47]	★	★	★	0	0	0	0	★	Poor
Lewis et al [48]	★	★	★	★	★	0	★	★	Good
Linke et al [50]	★	★	★	★	★	★	★	★	Good
Ma et al [51]	★	★	★	0	★	★	0	★	Good
Maher et al [52]	★	★	★	0	★	0	0	★	Poor
Marcus et al [22]	★	★	★	★	★	0	★	★	Good
Rebar et al [54]	★	★	★	0	★★	0	★	0	Poor
Wanner et al [55]	★	★	★	0	★	0	★	0	Poor
Xian et al [56]	★	★	★	★	★	★	0	★	Good

[a]Quality score: Overall scores were given (good, fair, and poor). Good quality: 3 or 4 stars (★) in the selection domain AND 1 or 2 stars in the comparability domain and 2 or 3 stars in the outcome domain; Fair quality: 2 stars in the selection domain and 1 or 2 stars in the comparability domain and 2 or 3 stars in the outcome/exposure domain; poor quality: 0 or 1 star in the selection domain OR 0 stars in the comparability domain OR 0 or 1 stars in the outcome/exposure domain.

Table 2. Quality assessment (Newcastle-Ottawa Quality Assessment Scale criteria for cross-sectional studies).

Study	Selection				Comparability	Outcome		Quality score[a]
	Representativeness of the exposed cohort	Sample size	Comparability of nonrespondents	Ascertainment of the exposure	Statistical analysis design features	Assessment of outcome	Statistical test	
Bronner et al [40]	0	★	0	★	0	★★	0	Poor
Davies et al [41]	★	0	★	★	★★	★	0	Poor
Hansen et al [45]	★	★	★	★	0	★	0	Poor
Hoj et al [46]	★	0	0	★	★	★	★	Fair
Lieber et al [49]	★	0	★	★	★	★	★	Good
Marquet et al [53]	★	0	0	★	★	★★	★	Fair

[a]Quality score: Overall scores were given (good, fair, and poor). Good quality: 3 or 4 stars (★) in the selection domain AND 1 or 2 stars in the comparability domain and 2 or 3 stars in the outcome domain; fair quality: 2 stars in the selection domain and 1 or 2 stars in the comparability domain and 2 or 3 stars in the outcome/exposure domain; poor quality: 0 or 1 star in the selection domain OR 0 stars in the comparability domain OR 0 or 1 stars in the outcome/exposure domain.

Selection

Within the cohort studies (n=13), all studies scored highly in the selection domain. Representativeness of the sample was high, with all studies scoring a star for being either truly or somewhat representative of the average target population. The nonexposed cohort was drawn from the same community as the exposed cohort in all studies. The exposure (engagement) was usually measured using either objective measurement (eg, Google Analytics) or self-report.

Within cross-sectional studies (n=6), all but one study had somewhat representative or truly representative samples. Sample size calculations were often not provided (n=4). Nonresponse characteristics were not provided or poorly described in half of the studies. No studies used validated measurement tools; however, the tool was made available or well described in all studies.

Comparability

Cohort studies controlled for confounders in 10 of the 13 studies. However, only one study controlled for all 3 factors (ie, age, sex, and marital status) required to score two stars. Therefore, most studies scored one star. Three studies scored zero stars, as they used unadjusted analyses.

In the cross-sectional studies, only 2 studies used adjusted analyses. One further study controlled for age and scored an additional star.

Outcome

Within the cohort studies, four studies used an independent blind assessment or record linkage (eg, steps via a mobile phone). The remaining studies scored zero stars as they used self-reporting. Eight studies were followed up after a sufficient duration (3 months); therefore, they scored a star. Five studies had follow-up shorter than 3 months. The follow-up cohort rate was inadequate in 3 studies, as no description of differences in responders and nonresponders was provided, and less than 80% responded. The remaining 10 studies scored a star, as either more than 80% responded at follow-up, or there were no differences in responders and nonresponders.

Of the cross-sectional studies, most studies used self-reported physical activity measurements (n=4). Half of the studies were considered to have used appropriate and well-described statistical tests; the remaining studies did not describe or provide sufficient details (eg, measures of variance).

Objective 1

Describe the direction and strength of the association between engagement with DHIs and physical activity or sedentary behavior in adults.

Although we had planned two meta-analyses, one for each of the conceptually different forms of engagement (use and subjective experience) [20], we did not conduct a meta-analysis for subjective experience because of the considerable methodological heterogeneity among the studies and the small number of studies reporting this outcome (n=3) [36]. Therefore, for this objective, we focused on the direction and strength of the association between usage and physical activity. There were 18 studies reporting a usage outcome, of which 7 were excluded from the meta-analysis [22,24,45,46,52,54,56], as data were not available to allow the calculation of an effect estimate or a measure of variability of the effect estimate data were not available.

The results from the meta-analysis of usage associations (n=11 studies) are shown in Figure 2 [41-44,47-51,53,55]. The characteristics of each of the associations from the meta-analysis are included in Table 3. The pooled estimate of the standardized regression coefficient (0.08; 95% CI 0.01-0.14; *P*=.02; SD 0.11) indicated a small but significant positive relationship between engagement with a DHI and physical activity. Heterogeneity was high, with 77% of the variation in the point estimates explained by the between-study heterogeneity.

Figure 2. Meta-analysis results from 11 studies to assess the direction and strength of the relationship between engagement with a digital health intervention and physical activity using the Dersimonian and Laird method.

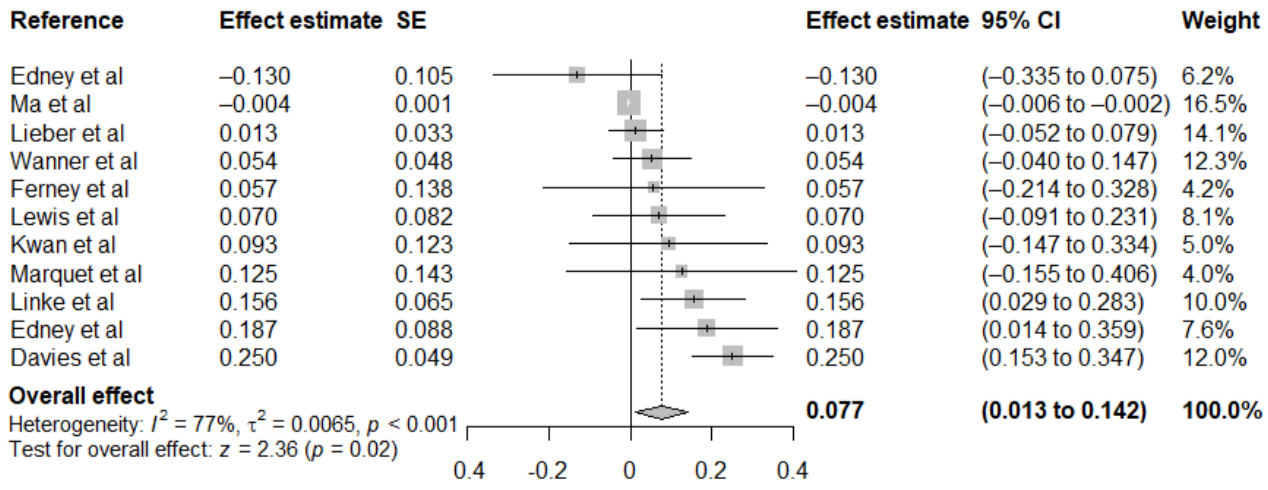


Table 3. Characteristics of studies (n=11) included in meta-analysis[a].

Study	DHI[b] type	Engagement measure	Physical activity measure
Edney et al [42]	Facebook	Activities completed	MVPA[c]
Ma et al [51]	Smartphone app	Time	Distance travelled
Lieber et al [49]	Web-based	Activities completed	MVPA
Wanner et al [55]	Web-based	Time	MVPA
Ferney et al [44]	Web-based	Logins	MVPA
Lewis et al [48]	Web-based	Activities completed	MVPA
Kwan et al [47]	Web-based	Logins	MVPA
Marquet et al [53]	Smartphone app	Time	Steps
Linke et al [50]	Web-based	Time	MVPA
Edney et al [43]	Smartphone app	Activities completed	MVPA
Davies et al [41]	Web-based	Activities completed	Steps

[a]A single association was selected for each study based on hierarchical criteria, see methods section. See Multimedia Appendix 5 for full details.

[b]DHI: digital health intervention.

[c]MVPA: moderate to vigorous physical activity.

The findings of the 7 studies that could not be included in the meta-analysis are summarized in Table 4. Four of the 7 studies reported an association consistent with the hypothesis that higher engagement (usage measures) is associated with a higher physical activity [22,45,52,56]. The remaining 3 studies had inconclusive findings [24,46,54]. Two studies had focused on the activities completed [24,56], one on time [54], and the remaining 4 on logins [22,45,46,52]. The inconclusive findings [24,46,54] were from the associations of different constructs of engagement (ie, activities completed, time, and logins).

Table 4. Characteristics and vote counting of studies (n=7) not included in meta-analysis.

Study	DHI[a] type	Engagement measure	Physical activity measure	Association type	Association	Direction[b]
Carr et al [24]	Web-based	Activities completed	Steps	Multiple linear regression	Text only: "No other predictors,...[including engagement]...significantly explained...change in physical activity."	0
Hansen et al [45]	Web-based	Logins	MVPA[c]	Kruskal-Wallis test	$P \leq .001$	+
Hoj et al [46]	Smartphone app	Logins	Physical activity score	Multiple regression	SE −0.01 (0.067)	0
Maher et al [52]	Web-based	Logins	MVPA	Generalized linear mixed models	($F_{1,41}$=3.06; P=.04)	+
Marcus et al [22]	Web-based	Logins	MVPA	Quantile regression	β=34.32 (95% CI 14.33 to 54.31)	+
Rebar et al [54]	Web-based	Time	MVPA	Linear mixed models	γ=0.51 (95% CI −1.77 to 2.72); P>.05	0
Xian et al [56]	Smartphone app	Activities completed	Steps	Ordinal least squares regression	Every 10,000 XP points gained were associated with 2134 additional steps per day (95% CI 1673 to 2595; P<.001; R^2=0.33])	+

[a]DHI: digital health intervention.

[b]A single association was selected for each study based on hierarchical criteria, see Methods section. Each study was summarized as either "+," "−," or "0." "+" was assigned to the studies in which the point estimate and CI supported the hypothesis that higher engagement is associated with higher physical activity or reduced sedentary behavior. "0" was assigned to studies in which the point estimate and CI had inconclusive findings. "−" was assigned to studies where the association point estimate and CI rejected the hypothesis. We assigned either "+" or "−" to the studies without point estimates or CIs that reported significant association findings. We assigned "0" to the studies without point estimates or CIs that reported nonsignificant findings. See Multimedia Appendix 5 and Methods section for full details.

[c]MVPA: moderate to vigorous physical activity.

Objective 2

Explore whether the direction of association between engagement and physical activity or sedentary behavior varies with the type of engagement (ie, subjective experience, activities completed, time, and logins).

Studies measured associations between physical activity and subjective experience (n=3) [40,46,48], activities completed (n=8) [24,41-43,48-50,56], time (n=5) [50,51,53-55], and logins (n=10) [22,24,44-48,50,52,54]. Therefore, 26 measures of association were included. The results of vote counting are summarized in Table 5. Overall, most associations (15 of 26) were in the hypothesized direction, stating that higher engagement is associated with higher physical activity. One association rejected the hypothesized direction, and the remaining 10 associations had inconclusive findings.

For the three domains of engagement, the direction consistently supported the hypothesis: subjective experience (2 of 3) [46,48], activities completed (5 of 8) [43,48-50,56], and logins (6 of 10) [22,45,48,50,52,54]. However, for time (n=5 associations), the findings did not support a positive association consistently, 2 studies had inconclusive findings [54,55], and one association rejected the hypothesized direction [51].

The 3 studies that described an association between subjective experience and physical activity used different measures to assess subjective experience. Bronner et al [40] used an Exergame Questionnaire that included questions similar to previously validated questions to assess subjective experience engagement in video games. The Exergame Questionnaire contained separate sections for engagement, game flow, and usability. The length of the questionnaire was not clear. Hoj et al [46] devised a five-question subjective experience questionnaire and constructed a composite score from it. Finally, Lewis et al [48] used the five-item Website Quality Questionnaire and constructed a composite score.

Table 5. Summary of associations included in vote counting.

Study	Engagement measure	DHI[a] type	Association type	Association	Direction[b]
Bronner et al [40]	Subjective experience	Exergame	Pearson's correlation	3 Associations: • ρ=0.61 • ρ=0.52 • Not reported	0
Hoj et al [46]	Subjective experience	Smartphone app	Multiple regression	SE 0.40 (0.074)	+
Lewis et al [48]	Subjective experience	Web-based	Quintile regression	t=2.32 (P≤.01)	+
Carr et al [24]	Activities completed	Web-based	Multiple regression	Not reported (nonsignificant)	0
Davies et al [41]	Activities completed	Web-based	Odds ratio	3 Associations: • OR[c] 2.80 (95% CI 1.45 to 5.40) • Not reported (nonsignificant) • Not reported (nonsignificant)	0
Edney et al [42]	Activities completed	Facebook group	Pearson's correlation	ρ=−0.13	0
Edney et al [43]	Activities completed	Smartphone app	Linear mixed models	$F_{1,272}$=4.5 (P=.04)	+
Lewis et al [48]	Activities completed	Web-based	Odds ratio	OR 1.29 (95% CI 1.14 to 1.47)	+
Lieber et al [49]	Activities completed	Web-based	Odds ratio	OR 1.05 (95% CI 1.01 to 1.09)	+
Linke et al [50]	Activities completed	Web-based	Generalized linear models	3 Associations: • β=2.85; SE: 1.38 (P=.04) • β=1.00; SE: 0.82 (P=.05) • β=3.49; SE: 1.28 (P=.01)	+
Xian et al [56]	Activities completed	Smartphone app	Ordinal least squares regression	Every 10,000 XP were associated with 2134 additional steps per day (95% CI 1673 to 2595; P<.001; R^2=0.33)	+
Linke et al [50]	Time	Web-based	Generalized linear models	β=0.48, SE: 0.20; P=.02	+
Ma et al [51]	Time	Smartphone app	Multi-level modelling	β=−0.005; P≤.001	−
Marquet et al [53]	Time	Smartphone app	ANCOVA	ρ=0.176; P<.05	+
Rebar et al [54]	Time	Web-based	Linear mixed models	2 associations: • γ=2.33 (95% CI 0.09 to 4.64); P<.05 • γ=0.51 (95% CI −1.77 to 2.72); P>.05	0
Wanner et al [55]	Time	Web-based	Linear regression	95% CI 0.58 (−0.43 to 1.59; P=.26)	0
Carr et al [24]	Logins	Web-based	Multiple regression	Not reported (nonsignificant)	0
Ferney et al [44]	Logins	Web-based	ANCOVA	4 Associations • P=.69 • P=.70 • P=.09 • P=.05	0
Hansen et al [45]	Logins	Web-based	Kruskal-Wallis test	P≤.001	+
Hoj et al [46]	Logins	Smartphone app	Multiple regression	SE −0.01 (0.067)	0
Kwan et al [47]	Logins	Web-based	ANOVA[d]	$F_{1,63}$=1.54, P=.22, n_p^2=0.03	0
Lewis et al [48]	Logins	Web-based	Quintile regression	t=3.39 (P≤.01)	+
Linke et al [50]	Logins	Web-based	Generalized linear models	Not reported (nonsignificant)	+

Study	Engagement measure	DHI[a] type	Association type	Association	Direction[b]
Maher et al [52]	Logins	Web-based	Generalized linear mixed models	$F_{1,41}$=3.06 (*P*=.04)	+
Marcus et al [22]	Logins	Web-based	Quantile regression	β=34.32 (95% CI 14.33 to 54.31)	+
Rebar et al [54]	Logins	Web-based	Linear mixed models	2 Associations: • γ=3.18 (95% CI 1.15 to 5.07); *P*<.05 • γ=2.04 (95% CI 0.29 to 3.84); *P*<.05	+

[a]DHI: digital health intervention.

[b]"+," "−," or "0" were assigned. "+" was assigned to studies where all associations within the particular engagement domain (subjective experience, activities completed, time and logins) where the point estimates and CIs supported the hypothesis that higher engagement is associated with higher physical activity or reduced sedentary behavior. "0" was assigned to the studies with inconclusive or mixed associations. "−" was assigned to the studies where all point estimates and CIs rejected the hypothesis that higher engagement is associated with higher physical activity or reduced sedentary behavior. See Multimedia Appendix 5 and Methods section for full details.

[c]OR: odds ratio.

[d]ANOVA: analysis of variance.

Discussion

Principal Findings

The findings of this review suggest that there is a positive relationship between engagement with a physical activity, both objective usage and subjective experience, and physical activity outcomes in adults. The strength of the relationship between DHI usage and physical activity based on a meta-analysis of 11 studies is weak (0.08, 95% CI 0.01-0.14). The direction of the association between physical activity and engagement was consistent across different measures of engagement, including two measures of usage (activities completed and logins) and subjective experience, but was less clear for the third measure of usage—time (ie, session duration). The majority of associations for subjective experience, activities completed, and logins were positive, whereas the remainder were inconclusive. There was a mixture of positive, inconclusive, and a negative association. No studies have examined the relationship between DHI engagement and sedentary behavior outcomes.

Findings in Context

This review updates by 10 years and expands on the review by Donkin et al [8], which identified 3 studies assessing the association between the usage of physical activity DHIs and physical activity outcomes [8]. In agreement with our review, Donkin et al [8] reported a consistent positive relationship between usage outcomes (eg, logins and activities completed) and DHIs targeting physical health (ie, psychological health, dietary behavior, physical activity, weight management, and smoking; 31/33 studies) [8]. Logins and activities completed were the most common engagement outcomes included in both the reviews. Donkin et al [8] did not include any studies with associations between time and physical health behavior, whereas our review contributes 5 studies [50,51,53-55]. Our findings for time were inconsistent, which aligns with the nonhealth studies exploring user engagement with internet-based news websites, which have found that time is not a reliable indicator of engagement [57].

We found a positive but weak relationship between DHI usage and physical activity. In contrast to the usage, it has been suggested that a clearer dose-response relationship exists between subjective experience engagement (eg, how captivating of attention a DHI is, the emotions a DHI elicits, and how interesting participants find a DHI) and effectiveness [20]. This further highlights the importance of defining the types of engagement outcomes as well as using multiple indicators of engagement when trying to understand the relationship between engagement variables and the effectiveness of DHIs [58,59].

This is the first review to examine the relationship between DHI subjective experience engagement and physical activity. Only 3 studies have reported associations between subjective experience and physical activity, with 2 of the associations in the hypothesized direction [40,46,48]. Each study used different self-reporting tools to assess different constructs of subjective experience (ie, website usefulness, app likeability, engagement, game-flow, and usability), which has previously been identified as an issue in the assessment of subjective experience engagement [60]. Such heterogeneity in constructs makes comparisons difficult, even though the direction of association is consistently positive. Future studies should focus on using consistent measures of subjective experience that are valid and reliable to enable comparisons between studies [61,62].

Further research into the relationship between subjective experience engagement and usage engagement is also warranted, as some qualitative studies suggest that usage is positively related to subjective experience [63,64]. For example, results from interviews with participants involved in an internet-based physical activity intervention reported that usage was positively influenced by subjective experience factors (eg, trust, reliability, and functionality of the program) [64]. Another study found that the sustained use of a Fitbit activity tracker was influenced by subjective experience–related factors (eg, empty batteries, broken trackers, and user experience) [63]. In other health behaviors, people who smoke and consume alcohol who wish to quit or cut down have suggested that the look, feel, app store quality rating, branding, and wording of the title are important

while choosing or not choosing to use an app [65]. Therefore, improving subjective experience could increase the strength of the positive relationship between usage and physical activity outcomes found in this review.

Strengths

A key strength of this review was the focus on two health behaviors (physical activity and sedentary behavior), reducing the heterogeneity and increasing the validity of findings [20]. This is the first review to examine the association between the subjective experience of DHI and health behavior. Including subjective experience and usage recognizes that engagement goes beyond usage, while considering attention, interest, and affect [20]. Excluding DHIs targeting individuals with a specific health condition reduces heterogeneity, as the context (including population) is known to influence engagement [20]. Another strength is the use of meta-analyses to examine the strength of the association between DHI usage and physical activity. Although it was not possible to conduct separate meta-analyses within engagement constructs, the use of vote-counting methods to assess the direction of association is a recommended method when meta-analyses are not possible [39].

Limitations

This review should be interpreted in the context of its limitations. The first limitation was that we included many different study types that produced large heterogeneity in the included studies. It means that we had to transform effect estimates to a common effect estimate and combine standardized effects, making interpretation of the results difficult. Furthermore, 8 studies were excluded from our meta-analysis because they provided insufficient information to be included in the meta-analysis. For vote counting, where studies did not report point estimates and CIs, we had to rely on the wording provided by the authors to infer whether the results supported our hypothesis. Second, the analysis of the association between engagement outcomes and physical activity outcomes in all 19 studies was a secondary analysis for these studies. Such analyses were often not well described. A further limitation was the inclusion of all recruitment types (ie, ecological and nonecological) within the same meta-analysis and vote-counting (ie, ecological and nonecological). It is known that ecological recruitment methods lead to higher attrition and lower engagement [31]. For example, Wanner et al [55] and Vandelanotte et al [31] highlight that spontaneous users (ecological) report a much lower engagement and higher dropout, whereas those that remain engaged become as active as those in the randomized groups (nonecological groups), possibly due to differing motivations [31,55]. In addition, although our search methods were rigorous, it is possible that expanding the search databases to include the ACM Digital Library may have identified additional studies from the human-computer interaction literature. Finally, given the lack of quantitative studies on subjective experience, perhaps owing to subjective experience being more often measured qualitatively [20], we encourage future reviews to explore the relationship between engagement and physical activity and sedentary behavior in qualitative studies.

Conclusions

A weak but consistent positive relationship exists between engagement with a physical activity DHI and physical activity outcomes. This is consistent across 2 of the 3 indicators of usage engagement that we examined, and subjective experience engagement; however, there are weak effect sizes. A further exploration of the relationship between engagement and physical activity using valid and reliable measurement tools is warranted, given the heterogeneity in measurement tools. Additional focus should be directed at DHI subjective experience (ie, attention, interest, and affect) by using consistent methodology to explore its relationship with the usage of DHIs and health behavior outcomes. Given the absence of studies, further research examining the association between DHIs and the impact on sedentary behavior is also warranted.

Acknowledgments

The authors convey sincere thanks to Debbie Booth, who was the Senior Librarian, for conducting the searches across scientific databases.

Authors' Contributions

MM drafted the manuscript. MM, TD, and LW developed the concept for the review. MM, TD, and JB screened all abstracts. AG, KR, MM, and TD screened all full texts. MM, TD, and AH extracted all data. MM, PM, and RS conducted the quality assessment. All authors contributed to the methods for the review, multiple versions of the manuscript, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 51 KB - jmir_v23i2e23180_app1.docx]

Multimedia Appendix 2

Newcastle Ottawa Scale for cohort studies.
[DOCX File , 16 KB - jmir_v23i2e23180_app2.docx]

Multimedia Appendix 3
Newcastle Ottawa Scale for cross-sectional studies.
[DOCX File , 16 KB - jmir_v23i2e23180_app3.docx]

Multimedia Appendix 4
Data transformations.
[DOCX File , 38 KB - jmir_v23i2e23180_app4.docx]

Multimedia Appendix 5
Supplementary table of characteristics of included studies.
[DOCX File , 77 KB - jmir_v23i2e23180_app5.docx]

References

1. Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. BMJ 2019 Aug 21:l4570 [FREE Full text] [doi: 10.1136/bmj.l4570]
2. Biswas A, Oh PI, Faulkner GE, Bajaj RR, Silver MA, Mitchell MS, et al. Sedentary Time and Its Association With Risk for Disease Incidence, Mortality, and Hospitalization in Adults. Ann Intern Med 2015 Jan 20;162(2):123 [FREE Full text] [doi: 10.7326/m14-1651]
3. Pandey A, Salahuddin U, Garg S, Ayers C, Kulinski J, Anand V, et al. Continuous Dose-Response Association Between Sedentary Time and Risk for Cardiovascular Disease. JAMA Cardiol 2016 Aug 01;1(5):575 [FREE Full text] [doi: 10.1001/jamacardio.2016.1567]
4. World Health Organization. Global action plan on physical activity 2018-2030: more active people for a healthier world. In: Noncommunicable diseases and their risk factors. Geneva: World Health Organization; 2018.
5. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017 Feb 2;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
6. International Telecommunication Union. ICT Facts and Figures 2016. Mobile network coverage and evolving technologies 2016:2016 [FREE Full text]
7. WHO. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. - 2016:1-144 [FREE Full text] [doi: 10.1017/cbo9780511712074.003]
8. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
9. World HO. WHO guideline: recommendations on digital interventions for health system strengthening. In: Sexual and reproductive health. Geneva: World Health Organization; 2019.
10. Müller AM, Maher CA, Vandelanotte C, Hingle M, Middelweerd A, Lopez ML, et al. Physical Activity, Sedentary Behavior, and Diet-Related eHealth and mHealth Research: Bibliometric Analysis. J Med Internet Res 2018 Apr 18;20(4):e122 [FREE Full text] [doi: 10.2196/jmir.8954] [Medline: 29669703]
11. WHO. mHealth: new horizons for health through mobile technologies. In: Global Observatory for eHealth series - Volume 3. Geneva: World Health Organanization; 2011.
12. Müller AM, Alley S, Schoeppe S, Vandelanotte C. Int J Behav Nutr Phys Act 2016 Oct 10;13(1):109 [FREE Full text] [doi: 10.1186/s12966-016-0434-2] [Medline: 27724911]
13. Maher C, Lewis L, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
14. Davies C, Spence J, Vandelanotte C, Caperchione C, Mummery W. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
15. Jahangiry L, Farhangi M, Shab-Bidar S, Rezaei F, Pashaei T. Web-based physical activity interventions: a systematic review and meta-analysis of randomized controlled trials. Public Health 2017;152:a.
16. Hou S, Charlery SR, Roberson K. Systematic literature review of Internet interventions across health behaviors. Health Psychol Behav Med 2014 Jan 01;2(1):455-481 [FREE Full text] [doi: 10.1080/21642850.2014.895368] [Medline: 25750795]

17. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can Smartphone Apps Increase Physical Activity? Systematic Review and Meta-Analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
18. Bort-Roig J, Gilson N, Puig-Ribera A, Contreras R, Trost S. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
19. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
20. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
21. McKay H, King D, Eakin E, Seeley J, Glasgow R. The diabetes network internet-based physical activity intervention: a randomized pilot study. Diabetes Care 2001 Aug;24(8):1328-1334. [doi: 10.2337/diacare.24.8.1328] [Medline: 11473065]
22. Marcus B, Lewis B, Williams D, Dunsiger S, Jakicic J, Whiteley J. A Comparison of Internet and Print-Based Physical Activity Interventions. JAMA Internal Medicine 2007;167(9):944-949. [Medline: 17502536]
23. Carr LJ, Bartee RT, Dorozynski CM, Broomfield JF, Smith ML, Smith DT. Eight-month follow-up of physical activity and central adiposity: results from an Internet-delivered randomized control trial intervention. J Phys Act Health 2009 Jul;6(4):444-455 [FREE Full text] [doi: 10.1123/jpah.6.4.444] [Medline: 19842458]
24. Carr LJ, Bartee RT, Dorozynski C, Broomfield JF, Smith ML, Smith DT. Internet-delivered behavior change program increases physical activity and improves cardiometabolic disease risk factors in sedentary adults: results of a randomized controlled trial. Prev Med 2008 May;46(5):431-438. [doi: 10.1016/j.ypmed.2007.12.005] [Medline: 18207228]
25. Short C, Rebar A, Plotnikoff R, Vandelanotte C. Designing engaging online behaviour change interventions a proposed model of user engagement. Netherlands urpean Health Psychology Society 2015:E [FREE Full text]
26. Moola S, Munn Z, Sears K, Sfetcu R, Currie M, Lisy K. Conducting systematic reviews of association (etiology): The Joanna Briggs Institute's approach. International journal of evidence-based healthcare 2015;13(3):163-169. [Medline: 26262566]
27. Tessa D, Mclaughlin M, Yoong S, Wyse R, Sutherland R, Hollis J, et al. Associations between digital health intervention engagement and dietary intake: a systematic review. 2018 Oct:2189 [FREE Full text] [doi: 10.13140/RG.2.2.32035.71203]
28. Dobbins M, Husson H, DeCorby K, LaRocca R. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. The Cochrane database of systematic reviews(2):Cd00 2013:7651 [FREE Full text]
29. Alkhaldi G, Hamilton F, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. Journal of medical Internet research 2016;18(1):a [FREE Full text]
30. Milward J, Drummond C, Fincham-Campbell S, Deluca P. What makes online substance-use interventions engaging? A systematic review and narrative synthesis. Digit Health 2018;4:2055207617743354 [FREE Full text] [doi: 10.1177/2055207617743354] [Medline: 29942622]
31. Vandelanotte C, Duncan MJ, Kolt GS, Caperchione CM, Savage TN, Van Itallie A, et al. More real-world trials are needed to establish if web-based physical activity interventions are effective. Br J Sports Med 2019 Dec;53(24):1553-1554. [doi: 10.1136/bjsports-2018-099437] [Medline: 29970409]
32. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
33. Wolfenden L, Goldman S, Stacey F, Grady A, Kingsland M, Williams C, et al. Strategies to improve the implementation of workplace-based policies or practices targeting tobacco, alcohol, diet, physical activity and obesity. Cochrane Database Syst Rev 2018 Nov 14;11:CD012439 [FREE Full text] [doi: 10.1002/14651858.CD012439.pub2] [Medline: 30480770]
34. Wells G, Shea B, O'Connell J. The Newcastle-Ottawa Scale (NOS) for assessing the quality of non randomised studies in meta-analyses. Ohrica. 2014. URL: https://www.researchgate.net/publication/288802810_The_Newcastle-Ottawa_Scale_NOS_for_Assessing_The_Quality_of_Nonrandomised_Studies_in_Meta-analyses [accessed 2020-12-23]
35. Modesti P, Reboldi G, Cappuccio F, Agyemang C, Remuzzi G, Rapi S, ESH Working Group on CV Risk in Low Resource Settings. Panethnic Differences in Blood Pressure in Europe: A Systematic Review and Meta-Analysis. PLoS One 2016;11(1):e0147601 [FREE Full text] [doi: 10.1371/journal.pone.0147601] [Medline: 26808317]
36. Michael BLJ. When Does it Make Sense to Perform a Meta‐Analysis? Chapter 40 2009:2020 [FREE Full text]
37. Nieminen P, Lehtiniemi H, Vähäkangas K, Huusko A, Rautio A. Standardised regression coefficient as an effect size index in summarising findings in epidemiological studies. Epidemiology, Biostatistics and Public Health 2013;10:- [FREE Full text] [doi: 10.2427/8854]
38. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/ [accessed 2020-12-23]

39. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020;368:- [FREE Full text]
40. Bronner S, Pinsker R, Adam NJ. Physiological and psychophysiological responses in experienced players while playing different dance exer-games. Computers in Human Behavior 2015;51:34-41 [FREE Full text]
41. Davies C, Corry K, Van Itallie A, Vandelanotte C, Caperchione C, Mummery W. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. J Med Internet Res 2012 Jan 11;14(1):e4 [FREE Full text] [doi: 10.2196/jmir.1792] [Medline: 22260810]
42. Edney S, Looyestyn J, Ryan J, Kernot J, Maher C. Posts, pics, or polls? Which post type generates the greatest engagement in a Facebook physical activity intervention? Transl Behav Med 2018 Nov 21;8(6):953-957. [doi: 10.1093/tbm/iby006] [Medline: 29660093]
43. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User Engagement and Attrition in an App-Based Physical Activity Intervention: Secondary Analysis of a Randomized Controlled Trial. J Med Internet Res 2019 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
44. Ferney SL, Marshall AL, Eakin EG, Owen N. Randomized trial of a neighborhood environment-focused physical activity website intervention. Prev Med 2009 Feb;48(2):144-150. [doi: 10.1016/j.ypmed.2008.10.022] [Medline: 19028519]
45. Hansen A, Grønbæk M, Helge J, Severin M, Curtis T, Tolstrup J. Effect of a Web-Based Intervention to Promote Physical Activity and Improve Health Among Physically Inactive Adults: A Population-Based Randomized Controlled Trial. J Med Internet Res 2012;14(5):A.
46. Hoj T, Covey E, Jones A, Haines A, Hall P, Crookston B. How Do Apps Work? An Analysis of Physical Activity App Users? Perceptions of Behavior Change Mechanisms. JMIR Mhealth Uhealth 2017;5(8):e114. [Medline: 28778846]
47. Kwan M, Faulkner G, Bray S. Evaluation of active transition, a website-delivered physical activity intervention for university students: pilot study. JMIR Res Protoc 2013 Apr 29;2(1):e16 [FREE Full text] [doi: 10.2196/resprot.2099] [Medline: 23649858]
48. Lewis B, Williams D, Dunsiger S, Sciamanna C, Whiteley J, Napolitano M, et al. User attitudes towards physical activity websites in a randomized controlled trial. Prev Med 2008 Nov;47(5):508-513 [FREE Full text] [doi: 10.1016/j.ypmed.2008.07.020] [Medline: 18773915]
49. Lieber SB, Redberg RF, Blumenthal RS, Gandhi A, Robb KJ, Mora S. A national interactive web-based physical activity intervention in women, evaluation of the american heart association choose to move program 2006-2007. Am J Cardiol 2012 Jun 15;109(12):1754-1760. [doi: 10.1016/j.amjcard.2012.02.017] [Medline: 22494850]
50. Linke SE, Dunsiger SI, Gans KM, Hartman SJ, Pekmezi D, Larsen BA, et al. Association Between Physical Activity Intervention Website Use and Physical Activity Levels Among Spanish-Speaking Latinas: Randomized Controlled Trial. J Med Internet Res 2019 Jul 24;21(7):e13063. [doi: 10.2196/13063]
51. Ma BD, Ng SL, Schwanen T, Zacharias J, Zhou M, Kawachi I, et al. Pokémon GO and Physical Activity in Asia: Multilevel Study. J Med Internet Res 2018 Jun 15;20(6):e217. [doi: 10.2196/jmir.9670]
52. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, De Bourdeaudhuij I, Thomas S, et al. A Web-Based, Social Networking Physical Activity Intervention for Insufficiently Active Adults Delivered via Facebook App: Randomized Controlled Trial. J Med Internet Res 2015 Jul 13;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
53. Marquet O, Alberico C, Hipp A. Pokémon GO and physical activity among college students. A study using Ecological Momentary Assessment. Computers in Human Behavior 2018 Apr;81:215-222. [doi: 10.1016/j.chb.2017.12.028]
54. Rebar A, Boles CW, W. Burton N, Duncan M, Short C, Happell B, et al. Healthy mind, healthy body: A randomized trial testing the efficacy of a computer-tailored vs. interactive web-based intervention for increasing physical activity and reducing depressive symptoms. Mental Health and Physical Activity 2016 Oct;11:29-37 [FREE Full text] [doi: 10.1016/j.mhpa.2016.08.001]
55. Wanner M, Martin-Diener E, Braun-Fahrländer C, Bauer G, Martin BW. Effectiveness of active-online, an individually tailored physical activity intervention, in a real-life setting: randomized controlled trial. J Med Internet Res 2009 Jul 28;11(3):e23. [doi: 10.2196/jmir.1179] [Medline: 19666456]
56. Xian Y, Xu H, Xu H, Liang L, Hernandez AF, Wang TY, et al. An Initial Evaluation of the Impact of Pokémon GO on Physical Activity. J Am Heart Assoc 2017 May 16;6(5):- [FREE Full text] [doi: 10.1161/JAHA.116.005341] [Medline: 28512111]
57. O'Brien H, Lebow M. Mixed-methods approach to measuring user experience in online news interactions. J Am Soc Inf Sci Tec 2013 Jun 19;64(8):1543-1556. [doi: 10.1002/asi.22871]
58. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
59. Cole-Lewis H, Ezeanochie N, Turgiss J. Understanding Health Behavior Technology Engagement: Pathway to Measuring Digital Behavior Change Interventions. JMIR Form Res 2019 Oct 10;3(4):e14052 [FREE Full text] [doi: 10.2196/14052] [Medline: 31603427]
60. Ng M, Firth J, Minen M, Torous J. User Engagement in Mental Health Apps: A Review of Measurement, Reporting, and Validity. Psychiatr Serv 2019 Jul 01;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]

61. Brooke J. SUS -- a quick and dirty usability scale. In: -. Oxfordshire, United Kingdom: Taylor & Francis; 1996:A.
62. O'Brien H, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. International Journal of Human-Computer Studies 2018 Apr;112:28-39. [doi: 10.1016/j.ijhcs.2018.01.004]
63. Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M. Determinants for Sustained Use of an Activity Tracker: Observational Study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164. [doi: 10.2196/mhealth.7311] [Medline: 29084709]
64. Bossen D, Buskermolen M, Veenhof C, de Bakker D, Dekker J. Adherence to a web-based physical activity intervention for patients with knee and/or hip osteoarthritis: a mixed method study. J Med Internet Res 2013 Oct 16;15(10):e223 [FREE Full text] [doi: 10.2196/jmir.2742] [Medline: 24132044]
65. Perski O, Blandford A, Ubhi H, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Feb 28;17(1):25 [FREE Full text] [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]

Abbreviations

DHI: digital health intervention
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register for Systematic Reviews

Edited by G Eysenbach; submitted 04.08.20; peer-reviewed by H O'Brien, C Vandelanotte; comments to author 11.09.20; revised version received 18.11.20; accepted 07.12.20; published 19.02.21.

Please cite as:
Mclaughlin M, Delaney T, Hall A, Byaruhanga J, Mackie P, Grady A, Reilly K, Campbell E, Sutherland R, Wiggers J, Wolfenden L Associations Between Digital Health Intervention Engagement, Physical Activity, and Sedentary Behavior: Systematic Review and Meta-analysis
J Med Internet Res 2021;23(2):e23180
URL: http://www.jmir.org/2021/2/e23180/
doi:10.2196/23180
PMID:33605897

Original Paper

Virtual and Augmented Reality Applications in Medicine: Analysis of the Scientific Literature

Andy Wai Kan Yeung[1,2], PhD; Anela Tosevska[2,3], PhD; Elisabeth Klager[2], MSc; Fabian Eibensteiner[2,4], MD; Daniel Laxar[2]; Jivko Stoyanov[5], PhD; Marija Glisic[5,6], PhD; Sebastian Zeiner[7], MD; Stefan Tino Kulnik[8], PhD; Rik Crutzen[8,9], PhD; Oliver Kimberger[2,7], MD, PD, PhD; Maria Kletecka-Pulker[2,10], PhD; Atanas G Atanasov[2,11,12,13], PD, PhD; Harald Willschke[2,7], MD, PhD

[1]Oral and Maxillofacial Radiology, Applied Oral Sciences and Community Dental Care, Faculty of Dentistry, The University of Hong Kong, Hong Kong, China

[2]Ludwig Boltzmann Institute for Digital Health and Patient Safety, Medical University of Vienna, Vienna, Austria

[3]Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA, United States

[4]Division of Pediatric Nephrology and Gastroenterology, Department of Pediatrics and Adolescent Medicine, Comprehensive Center for Pediatrics, Medical University of Vienna, Vienna, Austria

[5]Swiss Paraplegic Research, Nottwil, Switzerland

[6]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

[7]Department of Anaesthesia, Intensive Care Medicine and Pain Medicine, Medical University Vienna, Vienna, Austria

[8]Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria

[9]Department of Health Promotion, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, Netherlands

[10]Institute for Ethics and Law in Medicine, University of Vienna, Vienna, Austria

[11]Institute of Genetics and Animal Biotechnology of the Polish Academy of Sciences, Jastrzebiec, Poland

[12]Institute of Neurobiology, Bulgarian Academy of Sciences, Sofia, Bulgaria

[13]Department of Pharmacognosy, University of Vienna, Vienna, Austria

Corresponding Author:
Atanas G Atanasov, PD, PhD
Ludwig Boltzmann Institute for Digital Health and Patient Safety
Medical University of Vienna
Spitalgasse 23
Vienna,
Austria
Phone: 43 664 1929 852
Email: atanas.atanasov@univie.ac.at

Abstract

Background: Virtual reality (VR) and augmented reality (AR) have recently become popular research themes. However, there are no published bibliometric reports that have analyzed the corresponding scientific literature in relation to the application of these technologies in medicine.

Objective: We used a bibliometric approach to identify and analyze the scientific literature on VR and AR research in medicine, revealing the popular research topics, key authors, scientific institutions, countries, and journals. We further aimed to capture and describe the themes and medical conditions most commonly investigated by VR and AR research.

Methods: The Web of Science electronic database was searched to identify relevant papers on VR research in medicine. Basic publication and citation data were acquired using the "Analyze" and "Create Citation Report" functions of the database. Complete bibliographic data were exported to VOSviewer and Bibliometrix, dedicated bibliometric software packages, for further analyses. Visualization maps were generated to illustrate the recurring keywords and words mentioned in the titles and abstracts.

Results: The analysis was based on data from 8399 papers. Major research themes were diagnostic and surgical procedures, as well as rehabilitation. Commonly studied medical conditions were pain, stroke, anxiety, depression, fear, cancer, and neurodegenerative disorders. Overall, contributions to the literature were globally distributed with heaviest contributions from the United States and United Kingdom. Studies from more clinically related research areas such as surgery, psychology, neurosciences, and rehabilitation had higher average numbers of citations than studies from computer sciences and engineering.

Conclusions: The conducted bibliometric analysis unequivocally reveals the versatile emerging applications of VR and AR in medicine. With the further maturation of the technology and improved accessibility in countries where VR and AR research is strong, we expect it to have a marked impact on clinical practice and in the life of patients.

(J Med Internet Res 2021;23(2):e25499) doi:10.2196/25499

KEYWORDS

virtual reality; augmented reality; mixed reality; bibliometric; surgical procedures; rehabilitation; neurodegenerative disorder; pain; stroke; medicine

Introduction

Virtual reality (VR) is a technology that immerses the user in a synthetic 3-dimensional (3D) environment via wearable screens in the form of VR headsets, while closely related augmented reality (AR) uses elements of VR and superimposes them on to the real-world environment in the form of a live video displayed on the screen of an electronic device [1]. VR is a concept that has been developing over the last 50 years, whereas AR is a relatively new concept. Both are aimed at providing an experience for the users that engages their visual and auditory senses by creating an illusion of the surroundings [2,3]. In recent years, and with the advancement of technology, both VR and AR systems have become more portable, more realistic, and better to navigate in real time, adding a sensory and sometimes olfactory element to the range of sensations [3]. Furthermore, head-mounted devices are becoming more accessible. Personalized VR and AR devices have already been on the market for several years and are constantly improving and developing. The user can interact with the virtual environment through hand-held devices such as joysticks or keyboards and more recently, using integrated body tracking technologies [2,3]. VR and AR can be very versatile, using different systems and setups as well as different content that can range from very immersive, dynamic, and interactive to nonimmersive and static. Immersion, presence, and interaction are 3 essential characteristics of VR and AR [2,4]. Immersion is dependent on the technology used; it can be a head-mounted device, concave or 3D projection, or video where the user is the protagonist. Presence and interaction are related to the individual perception of being connected to the environment and the ability to act within the environment and receive feedback and reaction.

VR and AR are digital technologies that allow automation and can be used in fields where repetitive tasks need to be performed and often perfected [2]. A common example for the use of VR and AR in the medical field is medical education and training, especially in surgery [5-7]. For example, using VR or AR in surgical procedure training allows the trainee to perform steps on a virtual patient or having patient information superimposed with reality [1,5,7]. There is some evidence that VR could be a useful tool in improving surgical skills and reducing surgical procedure errors [8]. Neurosurgery, representing a traditionally complex surgical area, has been positively affected by the development of virtual techniques [9]. In acute pain management, VR has been used as a distraction technique [10,11], and there are studies proposing a role of VR in chronic pain management by inducing neurophysiological changes beyond simple distraction [10,12,13]. Further interesting prospects are using VR technology for the treatment of "phantom limb pain" following amputation [14] or after spinal cord injury [15]. VR has been used in rehabilitation for improvement of upper limb function following stroke [4], with modest or no improvement over conventional physiotherapy. Some benefits could be exhibited, on the other hand, in improving the cognitive abilities of patients with stroke, in particular speech, attention, and memory [16]. VR can also benefit patients with mental health conditions such as anxiety, depression, substance abuse, or eating disorders [17] and has been used as a therapy in a number of phobias and posttraumatic stress disorder [18].

The benefits of using VR or AR over conventional therapy could be plentiful: It might allow for multiple repetitions of simple tasks in clinical practice in an immersive environment without the need for constant supervision by medical staff, which could considerably reduce the costs for training facilities and trained medical staff. Furthermore, and especially for immobile patients, head-mounted devices could be safely used in patients' homes, which could decrease the need for hospital visits. VR and AR experiences can be designed to be attractive and user-friendly, decreasing the attrition rate of patients and providing a more pleasurable environment. From a research perspective, the use of VR can facilitate data collection for monitoring of progress [19]. Using VR in surgical training could vastly reduce the possibility for surgical errors, leading to great improvement in patient safety [8].

While the number of studies on the use of virtual technologies in health care is growing, these studies tend to be small and heterogeneous and often lack proper controls [4]. Results from such studies are often inconclusive, and the benefits of virtual over conventional approaches in health care are difficult to determine. The mechanisms by which VR treatment provides pain relief, for example, are still debated [12]. Moreover, a long-term benefit of VR treatment, especially in chronic pain management, has yet to be established. In a clinical setting, establishing VR systems is still technically challenging and cost-prohibitive, and such technology often has lower levels of acceptance in the elderly population [2]. A comprehensive guideline for standardization of the use of these technologies in medicine is still lacking and warrants further consideration.

Bibliometrics is an analytical approach that generates an integrative view and quantitative parameter profiling of entire research fields or specific scientific application areas [20-22]. Previous work has focused on specific areas of VR application such as dementia and rehabilitation medicine [23,24]. For instance, highly cited papers describing VR application for autism spectrum disorder (ASD) focused their research mainly

on the improvement of social skills [23], whereas research in the context of dementia mainly focused on the application of VR as an assessment tool for spatial navigation, memory profile, memory deficit, and memory formation in patients with mild cognitive impairment [24]. As VR is becoming a more prevalent topic in health care and medicine, we aimed to conduct a bibliometric analysis of the current literature, to discover trends and topics explored in VR applications in medicine and quantitatively evaluate the available literature. To the best of our knowledge, this is the first total-scale bibliometric analysis examining overall VR and AR applications in medicine in the scientific literature.

Methods

Data Source and Search Strategy

The Web of Science (WoS) Core Collection database was searched on September 16, 2020 and queried with the following search string: TOPIC: ("virtual reality*" OR "augmented reality*" OR "mixed reality*" OR "computer-mediated reality*") AND TOPIC: (medic* OR illness* OR disease* OR health* OR pharma*). The search identified publications mentioning these words and their derivatives in the title, abstract, or keywords. No additional restrictions, such as publication type or language, were used. The "Analyze" and "Create Citation Report" functions of WoS were utilized for basic publication and citation counting. The full records of the resultant publications were exported to VOSviewer (version 1.6.15) and Bibliometrix (Version 3.0 operated under the web interface called Biblioshiny) for further bibliometric analyses.

VOSviewer was used to produce a term map showing phrases from titles and abstracts of the publications. For clarity, phrases occurring in at least 0.5% (42/8399) of the publications were included. Multiple appearances in a single publication counted as one. In the obtained map, the circle size represents the frequency of occurrence, whereas the color represents the citations per publication. The distance between 2 circles represents how 2 phrases co-occurred with each other in the publications. Meanwhile, a density map was produced to show author´s keywords in the publications. For clarity, keywords occurring in at least 0.1% (9/8399) of the publications were included. A keyword or cluster of keywords with higher frequency counts formed a red region, and those with lower frequency counts formed a yellow region.

Utilizing the software Biblioshiny, the Trend Topics function was used to reveal trends in abstract words. Words were included if they occurred in at least 5 publications. Each year was limited to 5 words.

Statistical Analysis

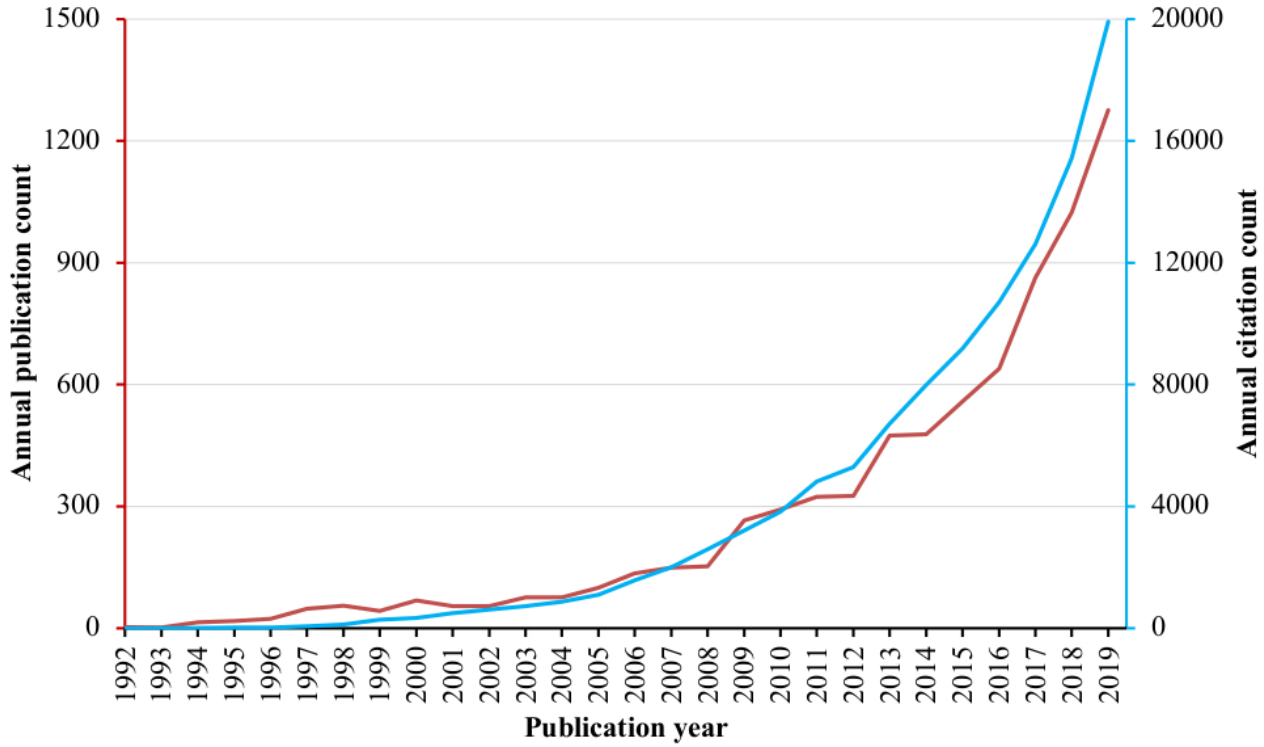
To analyze citation per publication (CPP) differences between original articles and reviews as well as between core and noncore journals, a 2-sample *t* test was conducted. In order to assess differences in CPP between different research areas (as designated by WoS), a one-way analysis of variance was utilized. Statistical tests were performed with SPSS (version 26.0; IBM Corp, Armonk, NY). Results were deemed significant if $P<.05$.

Results

Overall Literature Landscape

The literature search identified 8399 publications published from 1992 until 2020. The annual publication count reached 100 since 2005 and exceeded 1000 since 2018 (Figure 1). The year 2020 had 802 published papers until the date of the literature search (9/16/2020). There were 5297 original articles (5297/8399, 63.07%; CPP=17.4) and 902 reviews (902/8399, 10.74%; CPP=31.1). Therefore, the article-to-review ratio was 5.9:1, with reviews having a significantly higher CPP than original articles ($P<.001$). Other major document types were proceedings papers (1908/8399, 22.72%; CPP=2.0; WoS tagged 2213, of which 305 were simultaneously tagged as original articles), editorial materials (135/8399, 1.61%; CPP=13.7), and meeting abstracts (130/8399, 1.55%; CPP=0.2). The analyzed papers were mainly published in English (8131/8399, 96.81%).

Figure 1. Annual publication and citation count of virtual reality research in medicine.



Most Productive Entities

The 10 most productive authors by number of publications are listed in Table 1. The most productive author was Professor Giuseppe Riva from General Psychology and Communication Psychology, Catholic University of Milan, Italy. He also had the highest H-index among the 10 most productive authors. Meanwhile, the author with the highest CPP was Professor Albert “Skip” Rizzo from Medical Virtual Reality, Institute for Creative Technologies, University of Southern California. From the M-index, it was noted that among these authors, the most impactful newcomer to the field was Dr Silvia Serino from MySpace Lab, Department of Clinical Neurosciences, University Hospital Lausanne, Switzerland.

The 10 most productive organizations are listed in Table 2; 5 of these were in the United States, 2 each in the United Kingdom and Italy, and 1 in Canada. The University of London was the most productive organization, and the University of Toronto had the highest CPP, whereas the earliest research in this field was published at Harvard University.

Table 1. The 10 most productive authors of virtual reality research in medicine (8399 articles).

Author	Number of publications, n (%)	Citations per publication (CPP)	H-index[a]	M-index (first year)[a,b]
Giuseppe Riva	114 (1.36)	18.5	27	1.29 (2000)
Brenda K. Wiederhold	63 (0.75)	15.0	15	0.71 (2000)
Nassir Navab	60 (0.71)	12.6	16	0.76 (2000)
Albert “Skip” Rizzo	50 (0.60)	32.1	20	0.95 (2000)
Cristina Botella	37 (0.44)	17.0	16	0.80 (2001)
Mariano Alcaniz	36 (0.44)	22.7	15	0.75 (2001)
Pietro Cipresso	34 (0.40)	12.2	11	1.10 (2011)
Andrea Gaggioli	34 (0.40)	17.2	10	0.50 (2001)
Silvia Serino	33 (0.39)	14.6	13	1.63 (2013)
Lars Konge	32 (0.38)	13.8	12	1.20 (2011)

[a]Calculated from the dataset.

[b]Calculated by dividing the H-index by the number of years since the first published paper (within the dataset) of the author.

Table 2. The 10 most productive organizations of virtual reality research in medicine (8399 articles).

Organization	Publications, n (%)	Citations per publication (CPP)	H-index[a]	M-index (first year)[a,b]
University of London	175 (2.08)	27.5	37	1.61 (1998)
University of California System	164 (1.95)	22.4	34	1.42 (1997)
Harvard University	140 (1.67)	24.6	30	1.03 (1992)
University of Toronto	132 (1.57)	42.5	34	1.55 (1999)
Imperial College London	124 (1.48)	27.8	35	1.30 (1994)
Istituto Auxologico Italiano	116 (1.38)	18.1	27	1.29 (2000)
University of Southern California	113 (1.35)	29.1	29	1.07 (1994)
Catholic University of the Sacred Heart	112 (1.33)	17.5	25	1.19 (2000)
State University System of Florida	107 (1.27)	25.2	22	0.88 (1996)
Pennsylvania Commonwealth System of Higher Education	103 (1.23)	19.6	25	1.09 (1998)

[a]Calculated from the dataset.

[b]Calculated by dividing the H-index by the number of years since the first published paper (within the dataset) of the author.

The 10 most productive countries are listed in Table 3. The United States had contributions to nearly 30% of the VR publications in medicine and had the highest CPP. The international collaboration rates of these countries were mostly around 20%-30%, with the United States having a lower rate at 12.3%.

The 10 most productive journals are listed in Table 4. Computer science, surgery, and psychology were the 3 major research areas of these productive journals. Among the list, computer science journals seemed to have much lower CPP compared to others. According to Bradford's law, core journals are defined as the most productive journals that collectively account for publishing one-third of all concerned articles [25]. By this definition, there were 85 core and 3565 noncore journals in this dataset. The CPP of core (14.8) and noncore (15.0) journals did not significantly differ (P=.845).

Table 3. The 10 most productive countries of virtual reality research in medicine (8399 articles).

Country	Publications, n (%)	Citations per publication (CPP)	SCP[a]	MCP[b] (% of MCP:MCP+SCP)[c]
United States	2457 (29.25)	23.0	1807	253 (12.3)
United Kingdom	707 (8.42)	21.9	415	137 (24.8)
Germany	663 (7.89)	12.0	396	104 (20.8)
Canada	609 (7.25)	20.7	341	125 (26.8)
Italy	559 (6.66)	12.5	311	128 (29.2)
China	466 (5.55)	6.7	411	117 (22.2)
Spain	436 (5.19)	11.7	274	55 (16.7)
Australia	429 (5.11)	19.1	228	83 (26.7)
France	355 (4.23)	12.5	206	50 (19.5)
Netherlands	324 (3.86)	19.4	164	57 (25.8)

[a]SCP: single-country publication.

[b]MCP: multiple-country publication.

[c]SCP and MCP were computed by Bibliometrix based on data from the corresponding author's country only. Hence, their summation did not equal the total number of publications of that country.

Table 4. The 10 most productive journals of virtual reality research in medicine (8399 articles).

Journal	Publications, n (%)	Citations per publication (CPP)	2019 Impact Factor	Research area (domain)[a]
Lecture Notes in Computer Science	207 (2.46)	3.2	0.402	Computer science
Cyberpsychology, Behavior, and Social Networking	119 (1.42)	29.3	2.258	Psychology
Proceedings of SPIE[b]	97 (1.15)	1.6	N/A[c]	Computer science, optics
Surgical Endoscopy and Other Interventional Techniques	96 (1.14)	33.7	3.149	Surgery
Annual Review of Cybertherapy and Telemedicine	92 (1.10)	1.7	N/A	Computer science
PLOS One	81 (0.96)	12.4	2.740	Science and technology
Journal of Neuroengineering and Rehabilitation	73 (0.87)	27.0	3.519	Engineering, neurosciences and neurology, rehabilitation
Studies in Health Technology and Informatics	73 (0.87)	5.8	N/A	Health care sciences and services, medical informatics
Journal of Surgical Education	69 (0.82)	14.7	2.220	Education and educational research, surgery
Frontiers in Psychology	48 (0.57)	7.7	2.067	Psychology

[a]Research area as assigned by Web of Science.

[b]SPIE: Society of Photo-Optical Instrumentation Engineers.

[c]N/A: not available.

When the research area (of journals) was examined, the top 6 areas in descending order of publication count were computer science (CS), engineering (En), neurosciences and neurology (Neuro), surgery (Surg), psychology (Psy), and rehabilitation (Rehab). The CPPs of these 6 groups were significantly (P<.001) different from each other. Post-hoc tests revealed that the CPPs of Surg (25.1, SD 40.0), Psy (19.0, SD 35.6), and Neuro (20.3, SD 44.1) were greater than those of En (8.5, SD 26.8) and CS (9.5, SD 76.0); Surg was greater than that of Rehab (12.3, SD 24.5); and Others (15.6, SD 57.5) was greater than those of CS and En. Since some journals can be assigned to multiple research areas, a mutual exclusion procedure was performed in reverse hierarchical order (ie, the Rehab group retained all relevant papers first, then the Psy group retained all relevant papers excluding those coassigned to the Rehab group, then the Surg group retained all relevant papers excluding those coassigned to the Rehab and Psy groups, and so on). Finally, the Others group consisted of papers not assigned to the other 6 groups. In summary, papers dealing with clinical topics of surgery, neurosciences, psychology, and rehabilitation generally had more citations than those dealing with topics of computer science and engineering. This is consistent to a previous study in which neurosciences and surgery were among the top scientific categories with high citations [26].

Popular Research Themes

VOSviewer was used to generate a term map that tagged the phrases mentioned in the titles and abstracts of the publications with their CPPs (Figure 2). Phrases located in the upper part of the figure generally had higher CPPs. They were mainly related to surgery, such as laparoscopic skill (95/8399, CPP=29.9), surgical skill (190/8399, CPP=35.9), surgical training (195/8399, CPP=29.1), and surgical simulation (103/8399, CPP=26.5). Bibliometrix was used to illustrate the temporal changes in the abstract words (Figure 3). The most common words in the 2000s were related to diagnostic procedures such as angiographic, echocardiographic, ureteroscopy, and colonoscopy. In the 2010s, the most common words were related to brain diseases and mental health disorders, such as schizophrenia, AD (Alzheimer's disease), and ASD. Table 5 lists the medical conditions mentioned in at least 0.5% (42/8399) of the titles and abstracts of publications of VR research in medicine. This list was enriched with neuropsychological (eg, anxiety, depression), neurophysiological (eg, pain), and neuropathological (eg, Parkinson, Alzheimer) conditions and diseases.

Figure 2. Term map showing phrases from publication titles and abstracts in virtual reality research in medicine. Phrases occurring in at least 0.5% (42/8399) of the publications were included. Multiple appearances in a single publication counted as one. Circle size represents the frequency of occurrence, color represents the citation per publication, and distance between 2 circles represent how 2 phrases co-occurred in the publications.

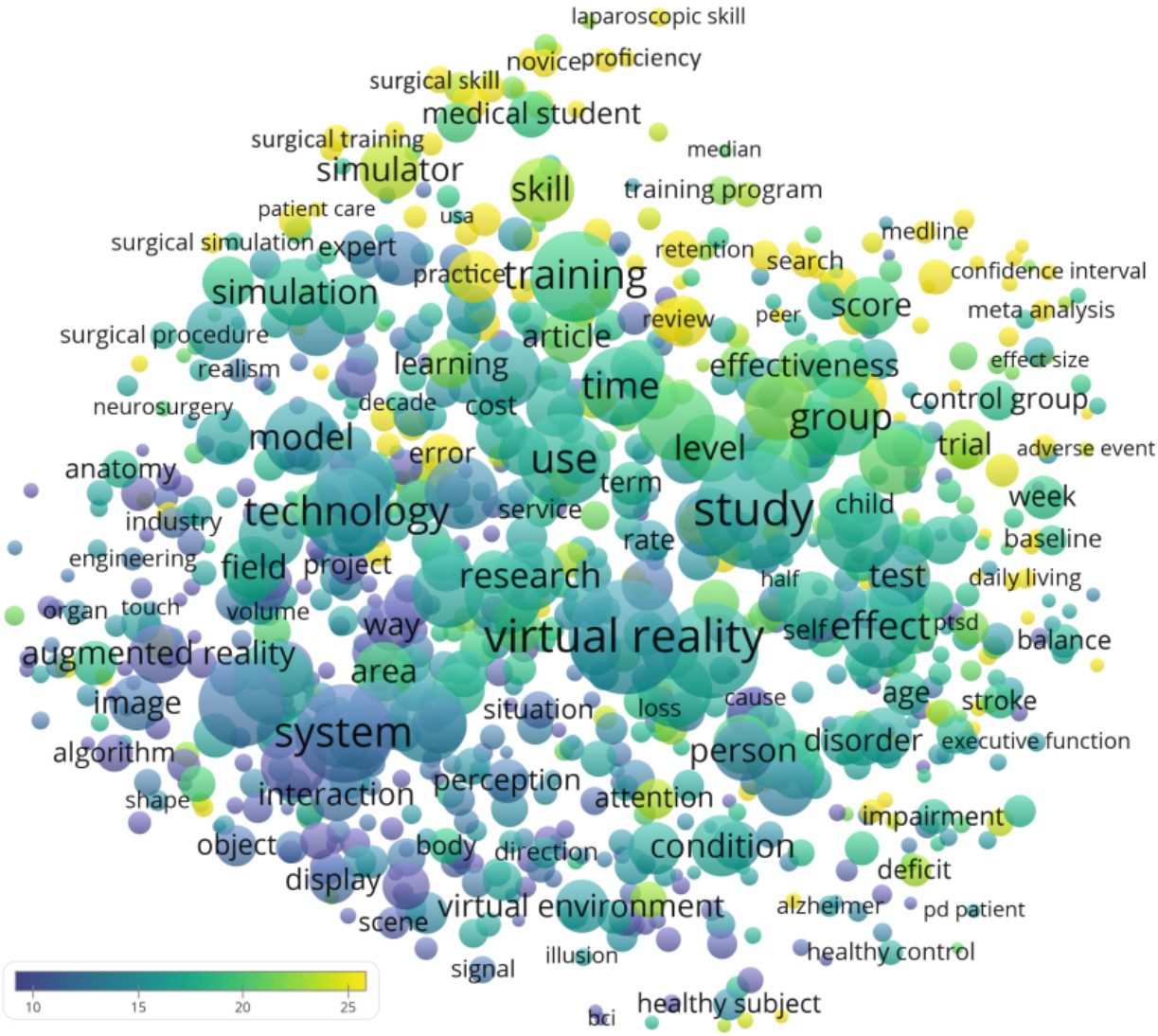


Figure 3. Trends in the abstract words, as determined using the Trend Topics function of Bibliometrix.

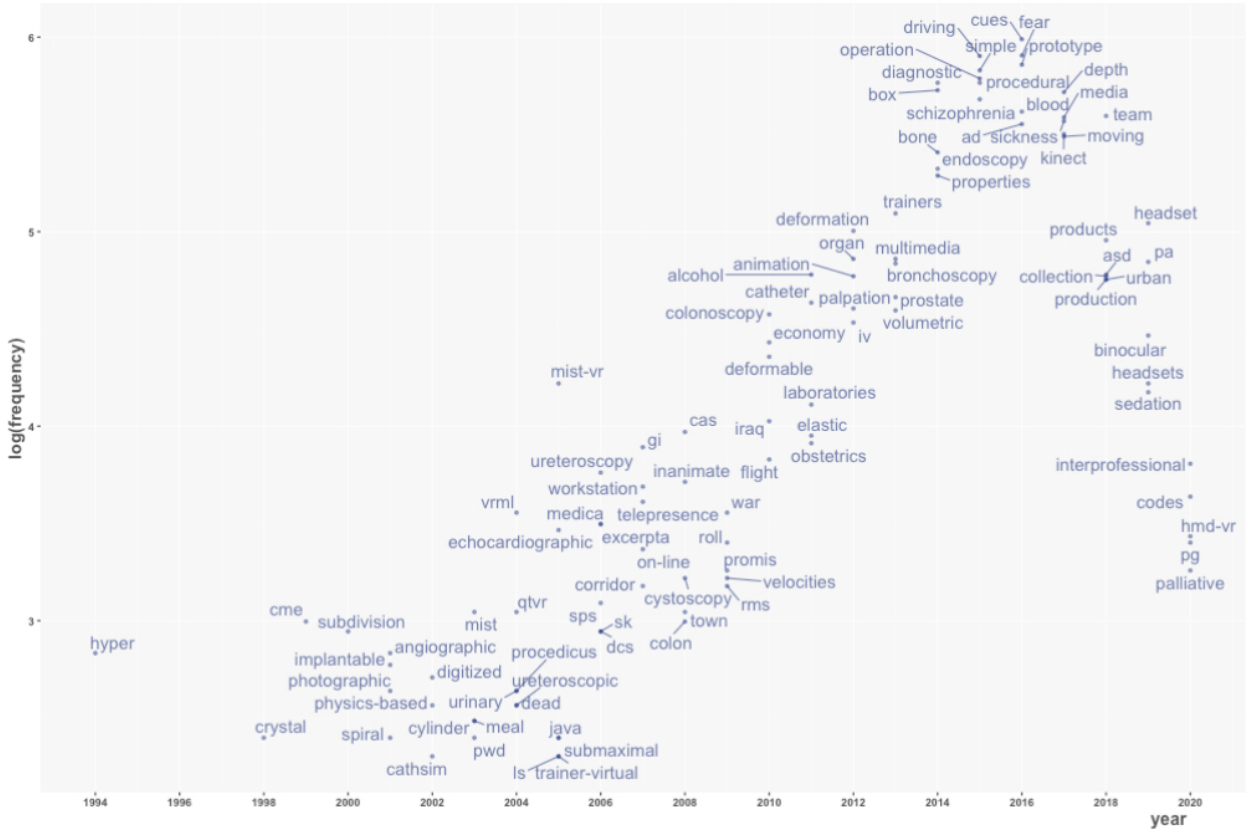


Table 5. Medical conditions mentioned in at least 0.5% (42/8399) of the publications of virtual reality research in medicine.

Medical condition	n	Citations per publication (CPP)
Pain	364	18.0
Stroke	319	18.3
Anxiety	316	21.2
Parkinson	258	17.6
Depression	148	18.3
Fear	143	24.7
Cancer	117	13.5
Alzheimer	113	15.1
Dementia	113	14.9
PTSD[a]	94	16.8
Mild cognitive impairment	90	19.7
Cerebral palsy	73	25.8
Traumatic brain injury	67	25.0
Multiple sclerosis	54	13.2
Psychosis	47	20.7
Obesity	46	11.5

[a]PTSD: posttraumatic stress disorder.

Author keywords also revealed that rehabilitation and simulation were the 2 major foci. As seen in Figure 4, the former was centered by words related to aging and neurodegenerative problems, such as Parkinson's disease (PD), stroke, dementia, gait, and cognition. The latter was centered by words related to surgery and education, such as surgical training, image-guided surgery, virtual patient, and robotic surgery. There were also other words related to emotion and perception, such as anxiety,

stress, and pain, which occupied subclusters with lower frequency densities. The 10 author keywords with the highest frequencies and CPPs are listed in Table 6.

The highest frequency keyword was simulation. Table 7 lists the 10 most cited papers that mentioned "simulation" in their keywords. Besides laparoscopic skills, the use of simulation in general medical education was also covered in terms of implementation issues, availability of products on the market, and recommendations.

Figure 4. Density map showing author keywords of the publications in virtual reality research in medicine. Keywords occurring in at least 0.1% (9/8399) of the publications were included. The keywords "virtual reality," "augmented reality," and "mixed reality" are not shown as they were the major search terms. A keyword or cluster of keywords with higher frequency counts forms a red region, and those with lower frequency counts form a yellow region.

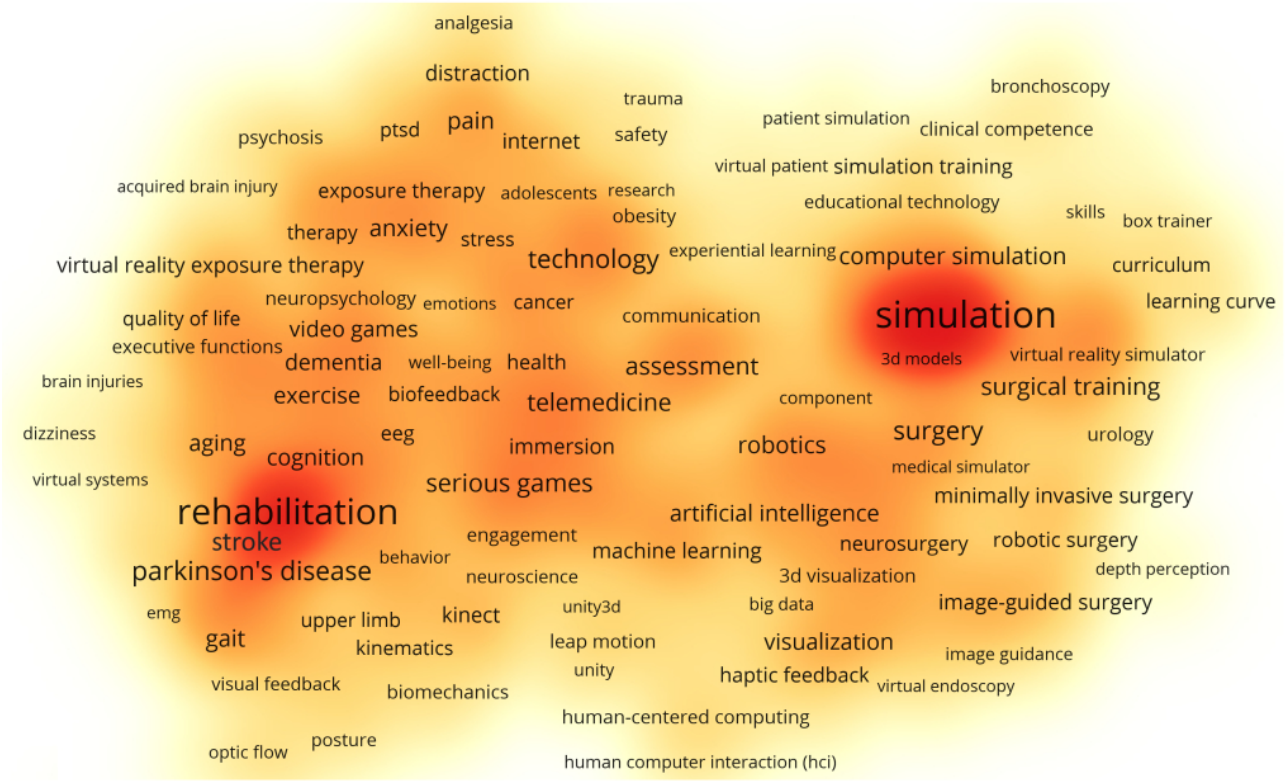


Table 6. The 10 author keywords with highest frequencies and citations per publication (CPPs; 8399 articles).

Keywords	Publications, n	CPP
Highest frequency		
Simulation	399	19.3
Rehabilitation	357	13.9
Education	263	15.9
Training	259	21.0
Stroke	211	11.8
Medical education	209	16.1
Laparoscopy	145	22.5
Parkinson's disease	126	17.1
Technology	109	9.5
Surgery	104	18.3
Highest CPP		
Optical flow	9	104.8
Hemiparesis	9	83.1
Patient simulation	13	78.0
Laparoscopic skills	9	74.4
Research	10	60.4
Education, medical	9	56.6
Hemiplegia	10	54.8
Laparoscopic	9	54.8
Motor control	21	46.3
Fear	9	45.2

Table 7. The 10 most cited papers with simulation as their keyword.

Authors	Title	Journal	Year	Citations, n
Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, Fried GM [27]	A global assessment tool for evaluation of intraoperative laparoscopic skills	American Journal of Surgery	2005	431
Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky III DA, Ramel S, Smith CD, Arvidsson D [28]	Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies	American Journal of Surgery	2007	367
Kneebone R [29]	Simulation in surgical training: educational issues and practical implications	Medical Education	2003	290
Kneebone RL, Scott W, Darzi A, Horrocks M [30]	Simulation and clinical practice: strengthening the relationship.	Medical Education	2004	215
Coles TR, Meglan D, John NW [31]	The role of haptics in medical training simulators: a survey of the state of the art	IEEE Transactions on Haptics	2010	203
Rosen KR [32]	The history of medical simulation	Journal of Critical Care	2008	188
Basdogan C, Ho CH, Srinivasan MA [33]	Virtual environments for medical training: graphical and haptic simulation of laparoscopic common bile duct exploration	IEEE/Asme Transactions on Mechatronics	2001	173
Hassfeld S, Mühling J [34]	Computer assisted oral and maxillofacial surgery–a review and an assessment of technology	International Journal of Oral and Maxillofacial Surgery	2001	164
Vozenilek J, Huff JS, Reznek M, Gordon JA [35]	See one, do one, teach one: advanced technology in medical education	Academic Emergency Medicine	2004	157
Zendejas B, Brydges R, Hamstra SJ, Cook DA [36]	State of the evidence on simulation-based training for laparoscopic surgery: a systematic review	Annals of Surgery	2013	154

The use of AR can be facilitated by many related technologies, for example, wearable tactile sensors [37] and skin-like electronics [38]. In surgery, AR can enhance the surgeon's vision by offering a virtual transparency of the patient [39]. When learning anatomy, AR can create an illusion that allows the display of internal anatomical structures on the body of the user [40]. These cases illustrate the versatility of AR.

Discussion

Principal Findings

This bibliometric analysis of 8399 publications on VR research in medicine revealed that the field began to develop in the 1990s, grew in the 2000s, and has been thriving in the 2010s in terms of both publications and citation counts. Original articles accounted for 63.1% of the literature. The article-to-review ratio was 5.9:1.

The literature had global contributions not only from North America and Europe but also from Asia and Oceania, implying that the application of VR in medicine has attracted worldwide attention. Diagnostic and surgical procedures as well as rehabilitation in neurodegenerative and mental health disorders were major research themes.

The use of VR simulators helped in diagnostic and surgical procedures, such as improving novice hospital residents' laparoscopic skills in terms of error and procedure time reduction [28]. It was also used to assess the suturing skills of different groups of operators [41]. In surgery, it was suggested that VR and simulation could be of value for 4 aspects, namely training and education, surgical planning, image guidance, and telesurgery [42].

Geriatrics has been exploring the use of VR tasks to assess and train episodic memory in the elderly population, by simulating various environments representative of daily life that cannot be physically replicated in the clinic or rehabilitation centers [43]. The use of VR might also help rehabilitation clinicians conduct telerehabilitation on a remote basis so that the patient (eg, after stroke) carries out exercises at home in a virtual environment and data are then transmitted to the clinician [44]. Moreover, VR could incorporate gamified elements so that the process could be more rewarding (eg, for encouraging patients with PD to do more remotely supervised aerobic exercise) [45]. Nonimmersive VR has also been used to add cognitive challenges and virtual obstacles to treadmill training for older adults, targeting attention, perception, and dual tasking during walking with the aim to reduce fall risk [46]. Since the existing studies were quite diverse, more studies should be conducted to optimize the implementation and evaluate the beneficiary effects in different population groups, so that a recommendation can be made for how to use VR in cognitive rehabilitation [47].

Meanwhile, the use of immersive VR was beneficial in managing a spectrum of emotional problems, such as the fear of heights [48], anxiety disorders [17], depression [49], and ASD [50]. Immersive VR could also reduce pain in patients, such as those with severe burn injuries during wound care [51] and pediatric patients during invasive medical procedures

[52,53]. Two functions of VR for this category are distraction [54] and creation of presence [55,56]. Whereas the former distracts the subject from the real-world situation, such as pain-inducing medical procedures, the latter enables the subject to "experience" designed scenarios to facilitate management strategies such as cognitive reappraisal or cognitive behavioral therapy.

To the authors' knowledge, no previous bibliometric analysis on VR research in the medicine literature overall has been published. A bibliometric analysis of VR research in general (including nonmedical areas) by Cipresso et al [2] similarly revealed the dominance by the United States, United Kingdom, Germany, and China in this research field [2]. Five authors in the top 10 list of this study were also in the top 10 list of Cipresso et al [2], namely Giuseppe Riva, Brenda K. Wiederhold, Albert "Skip" Rizzo, Cristina Botella, and Mariano Alcaniz. This implies that medicine might be among the most important scientific areas for applications of VR in general. The general VR research literature covered broader aspects apart from the research areas reported here and had large shares from mathematical and computational biology, radiology, and social sciences [2]. For VR research in ASD, the United States accounted for 51.2% of all publications, followed by the United Kingdom (16.5%), India, Spain, and China (3%-5% each) [57]. VR research in rehabilitation was similarly led by the United States (29.8%) and the United Kingdom (9.0%), followed by Italy (7.7%), Canada (7.4%), and Germany (5.1%) [58]. Meanwhile, the United Kingdom was most productive in VR research in dementia (43.2%), followed by the United States (39.5%), France (33.5%), Switzerland (28.6%), Germany, Greece, and Italy (23.2% each) [24].

A bibliometric analysis of artificial intelligence research in health and medicine similarly found that the United States was the top contributing country (30.8% versus 29.3% in this VR report) [59]. While China and Italy were the second and third most productive countries in artificial intelligence research in medicine, the United Kingdom did not enter the top 20 list. Here, for VR research in medicine, China and Italy were sixth and fifth on the list, whereas the United Kingdom ranked second. Both fields seemed to have different diseases of interest, but both shared stroke, cancer, AD, and PD [59].

Limitations

The current work searched one literature database only, implying that some papers could inevitably have been missed. However, as different databases record citation counts differently, it would be impossible to merge data from multiple databases. Moreover, publication and citation counts might not directly reflect the scientific quality of the papers, which might be influenced by various parameters including the sample size, study design, and standard of reporting. These would be best addressed by a systematic review with a focused scope, whereas the current work used bibliometric approach reports on the overall landscape of the literature in the research field.

Conclusions

Literature in the field of VR research in medicine represented over 8000 publications. The analyzed literature had global contributions with the heaviest contributions from the United States and United Kingdom. More clinically related research areas such as surgery, psychology, neurosciences, and rehabilitation had higher average numbers of citations than computer science and engineering. Diagnostic and surgical procedures and rehabilitation were major research themes. Medical conditions commonly investigated were pain, stroke, anxiety, depression, fear, cancer, and neurodegenerative disorders. The high potential and diversity of applications of VR and AR in medicine are already highly visible, and further improvements in these technologies are expected to both enhance their functionality and make them more accessible to patients, which will finally translate to significant therapeutic or preventive impact.

Conflicts of Interest

None declared.

References

1. Azuma RT. A Survey of Augmented Reality. Presence: Teleoperators & Virtual Environments 1997 Aug;6(4):355-385. [doi: 10.1162/pres.1997.6.4.355]
2. Cipresso P, Giglioli IAC, Raya MA, Riva G. The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature. Front Psychol 2018 Nov 6;9:2086 [FREE Full text] [doi: 10.3389/fpsyg.2018.02086] [Medline: 30459681]
3. Slater M, Sanchez-Vives MV. Enhancing Our Lives with Immersive Virtual Reality. Front. Robot. AI 2016 Dec 19;3:74. [doi: 10.3389/frobt.2016.00074]
4. Laver K, Lange B, George S, Deutsch J, Saposnik G, Crotty M. Virtual reality for stroke rehabilitation. Cochrane Database Syst Rev 2017 Nov 20;11:CD008349 [FREE Full text] [doi: 10.1002/14651858.CD008349.pub4] [Medline: 29156493]
5. Lewis T, Aggarwal R, Rajaretnam N, Grantcharov T, Darzi A. Training in surgical oncology - the role of VR simulation. Surg Oncol 2011 Sep;20(3):134-139. [doi: 10.1016/j.suronc.2011.04.005] [Medline: 21605972]
6. Samadbeik M, Yaaghobi D, Bastani P, Abhari S, Rezaee R, Garavand A. The Applications of Virtual Reality Technology in Medical Groups Teaching. J Adv Med Educ Prof 2018 Jul;6(3):123-129 [FREE Full text] [Medline: 30013996]
7. Seymour NE. VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. World J Surg 2008 Feb 3;32(2):182-188. [doi: 10.1007/s00268-007-9307-9] [Medline: 18060453]

8. Piromchai P, Avery A, Laopaiboon M, Kennedy G, O'Leary S. Virtual reality training for improving the skills needed for performing surgery of the ear, nose or throat. Cochrane Database Syst Rev 2015 Sep 09;9(9):CD010198. [doi: 10.1002/14651858.CD010198.pub2] [Medline: 26352008]
9. Fiani B, De Stefano F, Kondilis A, Covarrubias C, Reier L, Sarhadi K. Virtual Reality in Neurosurgery: "Can You See It?"-A Review of the Current Applications and Future Potential. World Neurosurg 2020 Sep;141:291-298. [doi: 10.1016/j.wneu.2020.06.066] [Medline: 32561486]
10. Ahmadpour N, Randall H, Choksi H, Gao A, Vaughan C, Poronnik P. Virtual Reality interventions for acute and chronic pain management. Int J Biochem Cell Biol 2019 Sep;114:105568. [doi: 10.1016/j.biocel.2019.105568] [Medline: 31306747]
11. Spiegel BM. Virtual medicine: how virtual reality is easing pain, calming nerves and improving health. Med J Aust 2018 Sep 17;209(6):245-247. [doi: 10.5694/mja17.00540] [Medline: 29764343]
12. Gupta A, Scott K, Dukewich M. Innovative Technology Using Virtual Reality in the Treatment of Pain: Does It Reduce Pain via Distraction, or Is There More to It? Pain Med 2018 Jan 01;19(1):151-159. [doi: 10.1093/pm/pnx109] [Medline: 29025113]
13. Pourmand A, Davis S, Marchak A, Whiteside T, Sikka N. Virtual Reality as a Clinical Tool for Pain Management. Curr Pain Headache Rep 2018 Jun 15;22(8):53. [doi: 10.1007/s11916-018-0708-2] [Medline: 29904806]
14. Dunn J, Yeo E, Moghaddampour P, Chau B, Humbert S. Virtual and augmented reality in the treatment of phantom limb pain: A literature review. NRE 2017 Jun 09;40(4):595-601. [doi: 10.3233/nre-171447]
15. Pozeg P, Palluel E, Ronchi R, Solcà M, Al-Khodairy A, Jordan X, et al. Virtual reality improves embodiment and neuropathic pain caused by spinal cord injury. Neurology 2017 Oct 06;89(18):1894-1903. [doi: 10.1212/wnl.0000000000004585]
16. Maggio M, Latella D, Maresca G, Sciarrone F, Manuli A, Naro A, et al. Virtual Reality and Cognitive Rehabilitation in People With Stroke: An Overview. J Neurosci Nurs 2019 Apr;51(2):101-105. [doi: 10.1097/JNN.0000000000000423] [Medline: 30649091]
17. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol. Med 2017 Mar 22;47(14):2393-2400. [doi: 10.1017/s003329171700040x]
18. Maples-Keller JL, Yasinski C, Manjin N, Rothbaum BO. Virtual Reality-Enhanced Extinction of Phobias and Post-Traumatic Stress. Neurotherapeutics 2017 Jul 16;14(3):554-563 [FREE Full text] [doi: 10.1007/s13311-017-0534-y] [Medline: 28512692]
19. Tieri G, Morone G, Paolucci S, Iosa M. Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies. Expert Rev Med Devices 2018 Feb 10;15(2):107-117. [doi: 10.1080/17434440.2018.1425613] [Medline: 29313388]
20. Yeung AWK, Souto EB, Durazzo A, Lucarini M, Novellino E, Tewari D, et al. Big impact of nanoparticles: analysis of the most cited nanopharmaceuticals and nanonutraceuticals research. Current Research in Biotechnology 2020 Nov;2:53-63. [doi: 10.1016/j.crbiot.2020.04.002]
21. Yeung AWK, Tzvetkov NT, Georgieva MG, Ognyanov IV, Kordos K, Jóźwik A, et al. Reactive Oxygen Species and Their Impact in Neurodegenerative Diseases: Literature Landscape Analysis. Antioxid Redox Signal 2021 Feb 10;34(5):402-420. [doi: 10.1089/ars.2019.7952] [Medline: 32030995]
22. Yeung AWK, Tzvetkov NT, Jóźwik A, Horbanczuk OK, Polgar T, Pieczynska MD, et al. Food toxicology: quantitative analysis of the research field literature. Int J Food Sci Nutr 2020 Feb 29;71(1):13-21. [doi: 10.1080/09637486.2019.1620184] [Medline: 31140340]
23. Fernández-Herrero J, Lorenzo-Lledó G, Carreres A. A bibliometric study on the use of virtual reality (VR) as an educational tool for high-functioning Autism Spectrum Disorder (ASD) children. In: Çetinkaya S, editor. Contemporary Perspective on Child Psychology and Education. London, England: IntechOpen Limited; 2017.
24. Sobral M, Pestana M. Virtual reality and dementia: A bibliometric analysis. The European Journal of Psychiatry 2020 Jul;34(3):120-131. [doi: 10.1016/j.ejpsy.2020.04.004]
25. Vickery B. Bradford's Law of Scattering. Journal of Documentation 1948 Apr;4(3):198-203. [doi: 10.1108/eb026133]
26. Patience GS, Patience CA, Blais B, Bertrand F. Citation analysis of scientific categories. Heliyon 2017 May;3(5):e00300. [doi: 10.1016/j.heliyon.2017.e00300]
27. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 2005 Jul;190(1):107-113. [doi: 10.1016/j.amjsurg.2005.04.004] [Medline: 15972181]
28. Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DA, et al. Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. Am J Surg 2007 Jun;193(6):797-804. [doi: 10.1016/j.amjsurg.2006.06.050] [Medline: 17512301]
29. Kneebone R. Simulation in surgical training: educational issues and practical implications. Med Educ 2003 Mar;37(3):267-277. [doi: 10.1046/j.1365-2923.2003.01440.x] [Medline: 12603766]
30. Kneebone RL, Scott W, Darzi A, Horrocks M. Simulation and clinical practice: strengthening the relationship. Med Educ 2004 Oct;38(10):1095-1102. [doi: 10.1111/j.1365-2929.2004.01959.x] [Medline: 15461655]
31. Coles TR, Meglan D, John NW. The Role of Haptics in Medical Training Simulators: A Survey of the State of the Art. IEEE Trans Haptics 2011;4(1):51-66. [doi: 10.1109/TOH.2010.19] [Medline: 26962955]

32. Rosen KR. The history of medical simulation. J Crit Care 2008 Jun;23(2):157-166. [doi: 10.1016/j.jcrc.2007.12.004] [Medline: 18538206]
33. Basdogan C, Ho C, Srinivasan M. Virtual environments for medical training: graphical and haptic simulation of laparoscopic common bile duct exploration. IEEE/ASME Trans. Mechatron 2011 Mar 15;6(3):269-285 [FREE Full text] [doi: 10.1109/3516.951365] [Medline: 21367734]
34. Hassfeld S, Mühling J. Computer assisted oral and maxillofacial surgery--a review and an assessment of technology. Int J Oral Maxillofac Surg 2001 Feb;30(1):2-13. [doi: 10.1054/ijom.2000.0024] [Medline: 11289616]
35. Vozenilek J, Huff JS, Reznek M, Gordon JA. See one, do one, teach one: advanced technology in medical education. Acad Emerg Med 2004 Nov;11(11):1149-1154 [FREE Full text] [doi: 10.1197/j.aem.2004.08.003] [Medline: 15528578]
36. Zendejas B, Brydges R, Hamstra SJ, Cook DA. State of the evidence on simulation-based training for laparoscopic surgery: a systematic review. Ann Surg 2013 Apr;257(4):586-593. [doi: 10.1097/SLA.0b013e318288c40b] [Medline: 23407298]
37. Yang T, Xie D, Li Z, Zhu H. Recent advances in wearable tactile sensors: Materials, sensing mechanisms, and device performance. Materials Science and Engineering: R: Reports 2017 May;115:1-37. [doi: 10.1016/j.mser.2017.02.001]
38. Wang S, Xu J, Wang W, Wang GN, Rastak R, Molina-Lopez F, et al. Skin electronics from scalable fabrication of an intrinsically stretchable transistor array. Nature 2018 Mar 01;555(7694):83-88. [doi: 10.1038/nature25494] [Medline: 29466334]
39. Nicolau S, Soler L, Mutter D, Marescaux J. Augmented reality in laparoscopic surgical oncology. Surg Oncol 2011 Sep;20(3):189-201. [doi: 10.1016/j.suronc.2011.07.002] [Medline: 21802281]
40. Estai M, Bunt S. Best teaching practices in anatomy education: A critical review. Ann Anat 2016 Nov;208:151-157. [doi: 10.1016/j.aanat.2016.02.010] [Medline: 26996541]
41. O'Toole RV, Playter R, Krummel T, Blank W, Cornelius N, Roberts W, et al. Measuring and developing suturing technique with a virtual reality surgical simulator. Journal of the American College of Surgeons 1999 Jul;189(1):114-127. [doi: 10.1016/s1072-7515(99)00076-9]
42. Gorman PJ, Meier AH, Krummel TM. Simulation and virtual reality in surgical education: real or unreal? Arch Surg 1999 Nov 01;134(11):1203-1208. [doi: 10.1001/archsurg.134.11.1203] [Medline: 10555634]
43. Corriveau Lecavalier N, Ouellet E, Boller B, Belleville S. Use of immersive virtual reality to assess episodic memory: A validation study in older adults. Neuropsychol Rehabil 2020 Apr 29;30(3):462-480. [doi: 10.1080/09602011.2018.1477684] [Medline: 29807474]
44. Laver K, Schoene D, Crotty M, George S, Lannin NA, Sherrington C. Telerehabilitation services for stroke. Cochrane Database Syst Rev 2013 Dec 16;1(12):CD010255 [FREE Full text] [doi: 10.1002/14651858.CD010255.pub2] [Medline: 24338496]
45. van der Kolk NM, de Vries NM, Kessels RPC, Joosten H, Zwinderman AH, Post B, et al. Effectiveness of home-based and remotely supervised aerobic exercise in Parkinson's disease: a double-blind, randomised controlled trial. The Lancet Neurology 2019 Nov;18(11):998-1008. [doi: 10.1016/s1474-4422(19)30285-6]
46. Mirelman A, Rochester L, Maidan I, Del Din S, Alcock L, Nieuwhof F, et al. Addition of a non-immersive virtual reality component to treadmill training to reduce fall risk in older adults (V-TIME): a randomised controlled trial. The Lancet 2016 Sep;388(10050):1170-1182. [doi: 10.1016/s0140-6736(16)31325-3]
47. Cicerone KD, Goldin Y, Ganci K, Rosenbaum A, Wethe JV, Langenbahn DM, et al. Evidence-Based Cognitive Rehabilitation: Systematic Review of the Literature From 2009 Through 2014. Arch Phys Med Rehabil 2019 Aug;100(8):1515-1533. [doi: 10.1016/j.apmr.2019.02.011] [Medline: 30926291]
48. Freeman D, Haselton P, Freeman J, Spanlang B, Kishore S, Albery E, et al. Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. The Lancet Psychiatry 2018 Aug;5(8):625-632. [doi: 10.1016/s2215-0366(18)30226-8]
49. Falconer CJ, Rovira A, King JA, Gilbert P, Antley A, Fearon P, et al. Embodying self-compassion within virtual reality and its effects on patients with depression. BJPsych Open 2016 Jan 02;2(1):74-80 [FREE Full text] [doi: 10.1192/bjpo.bp.115.002147] [Medline: 27703757]
50. Malinverni L, Mora-Guiard J, Padillo V, Valero L, Hervás A, Pares N. An inclusive design approach for developing video games for children with Autism Spectrum Disorder. Computers in Human Behavior 2017 Jun;71:535-549. [doi: 10.1016/j.chb.2016.01.018]
51. Hoffman HG, Chambers GT, Meyer WJ, Arceneaux LL, Russell WJ, Seibel EJ, et al. Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. Ann Behav Med 2011 Apr 25;41(2):183-191 [FREE Full text] [doi: 10.1007/s12160-010-9248-7] [Medline: 21264690]
52. Gold JI, Kim SH, Kant AJ, Joseph MH, Rizzo A. Effectiveness of virtual reality for pediatric pain distraction during i.v. placement. Cyberpsychol Behav 2006 Apr;9(2):207-212. [doi: 10.1089/cpb.2006.9.207] [Medline: 16640481]
53. Gershon J, Zimand E, Pickering M, Rothbaum BO, Hodges L. A pilot and feasibility study of virtual reality as a distraction for children with cancer. J Am Acad Child Adolesc Psychiatry 2004 Oct;43(10):1243-1249. [doi: 10.1097/01.chi.0000135621.23145.05] [Medline: 15381891]
54. Kenney MP, Milling LS. The effectiveness of virtual reality distraction for reducing pain: A meta-analysis. Psychology of Consciousness: Theory, Research, and Practice 2016;3(3):199-210. [doi: 10.1037/cns0000084]

55. David D, Matu S, David OA. New Directions in Virtual Reality-Based Therapy for Anxiety Disorders. International Journal of Cognitive Therapy 2013 Jun;6(2):114-137. [doi: 10.1521/ijct.2013.6.2.114]
56. Riva G, Mantovani F, Capideville CS, Preziosa A, Morganti F, Villani D, et al. Affective interactions using virtual reality: the link between presence and emotions. Cyberpsychol Behav 2007 Feb;10(1):45-56. [doi: 10.1089/cpb.2006.9993] [Medline: 17305448]
57. Fernández-Herrero J, Lorenzo-Lledó G, Carreres A. A bibliometric study on the use of virtual reality (VR) as an educational tool for high-functioning Autism Spectrum Disorder (ASD) children. Contemporary Perspective on Child Psychology and Education 2018:59-81. [doi: 10.5772/intechopen.71000]
58. Huang Y, Huang Q, Ali S, Zhai X, Bi X, Liu R. Rehabilitation using virtual reality technology: a bibliometric analysis, 1996–2015. Scientometrics 2016 Oct 05;109(3):1547-1559. [doi: 10.1007/s11192-016-2117-9]
59. Tran B, Vu G, Ha G, Vuong Q, Ho M, Vuong T, et al. Global Evolution of Research in Artificial Intelligence in Health and Medicine: A Bibliometric Study. J Clin Med 2019 Mar 14;8(3):360 [FREE Full text] [doi: 10.3390/jcm8030360] [Medline: 30875745]

Abbreviations

3D: 3-dimensional
AD: Alzheimer's disease
AR: augmented reality
ASD: autism spectrum disorder
CPP: citation per publication
CS: computer science
En: engineering
Neuro: neurosciences & neurology
PD: Parkinson's disease
Psy: psychology
Rehab: rehabilitation
Surg: surgery
VR: virtual reality
WoS: Web of Science

Edited by G Eysenbach; submitted 05.11.20; peer-reviewed by MS Uddin; comments to author 26.11.20; revised version received 08.12.20; accepted 16.01.21; published 10.02.21.

Please cite as:
Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Laxar D, Stoyanov J, Glisic M, Zeiner S, Kulnik ST, Crutzen R, Kimberger O, Kletecka-Pulker M, Atanasov AG, Willschke H
Virtual and Augmented Reality Applications in Medicine: Analysis of the Scientific Literature
J Med Internet Res 2021;23(2):e25499
URL: http://www.jmir.org/2021/2/e25499/
doi:10.2196/25499
PMID:33565986

Review

Virtual Reality or Augmented Reality as a Tool for Studying Bystander Behaviors in Interpersonal Violence: Scoping Review

Jia Xue[1,2], PhD; Ran Hu[1], MA, MSW; Wenzhao Zhang[2], MA; Yaxi Zhao[2], MA; Bolun Zhang[2], MI, MA; Nian Liu[2], BA; Sam-Chin Li[3], MA, MLS; Judith Logan[3], MA, MLIS

[1]Factor Inwentash Faculty of Social Work, University of Toronto, Toronto, ON, Canada

[2]Faculty of Information, University of Toronto, Toronto, ON, Canada

[3]University of Toronto Libraries, Toronto, ON, Canada

Corresponding Author:
Jia Xue, PhD
Factor Inwentash Faculty of Social Work
University of Toronto
246 Bloor Street West
Toronto, ON, M5S 1V4
Canada
Phone: 1 4169465429
Email: jia.xue@utoronto.ca

Abstract

Background: To provide participants with a more real and immersive intervening experience, virtual reality (VR) and/or augmented reality (AR) technologies have been integrated into some bystander intervention training programs and studies measuring bystander behaviors.

Objective: We focused on whether VR or AR can be used as a tool to enhance training bystanders. We reviewed the evidence from empirical studies that used VR and/or AR as a tool for examining bystander behaviors in the domain of interpersonal violence research.

Methods: Two librarians searched for articles in databases, including APA PsycInfo (Ovid), Criminal Justice Abstracts (EBSCO), Medline (Ovid), Applied Social Sciences Index & Abstracts (ProQuest), Sociological Abstracts (ProQuest), and Scopus till April 15, 2020. Studies focusing on bystander behaviors in conflict situations were included. All study types (except reviews) written in English in any discipline were included.

Results: The search resulted in 12,972 articles from six databases, and the articles were imported into Covidence. Eleven studies met the inclusion and exclusion criteria. All 11 articles examined the use of VR as a tool for studying bystander behaviors. Most of the studies were conducted in US young adults. The types of interpersonal violence were school bullying, dating violence, sexual violence/assault, and soccer-associated violence. VR technology was used as an observational measure and bystander intervention program. We evaluated the different uses of VR for bystander behaviors and noted a lack of empirical evidence for AR as a tool. We also discuss the empirical evidence regarding the design, effectiveness, and limitations of implementing VR as a tool in the reviewed studies.

Conclusions: The reviewed results have implications and recommendations for future research in designing and implementing VR/AR technology in the area of interpersonal violence. Future studies in this area may further contribute to the use of VR as an observational measure and explore the potential use of AR to study bystander behaviors.

(J Med Internet Res 2021;23(2):e25322) doi:10.2196/25322

KEYWORDS

virtual reality; augmented reality; bystander behaviors; interpersonal violence; violent incidents; people's responses; dating violence; sexual violence

Introduction

Interpersonal Violence and Bystander Intervention

According to the World Health Organization, interpersonal violence refers to "the intentional use of physical force or power, threatened or actual" against individuals who may be family members (including intimate partners), colleagues, acquaintances, or strangers [1]. Drawing on this definition, common forms of interpersonal violence include physical violence, verbal abuse, psychological violence, and sexual assault or harassment [2]. Given the multifaceted and detrimental impact of interpersonal violence on the victims (eg, health and mental health consequences [3-5]), developing effective strategies to prevent violence in the first place becomes crucial. Those who witness violent incidents or potential risks for violence, often known as bystanders, may play an influential role in intervening in the situation, such as providing support to the victims and reporting the incident [6,7].

Empirical studies [8] have focused on examining factors that may influence bystander intervening behaviors, such as bystander self-efficacy [9], skills to intervene [6], and awareness of signs of violence [10]. Using an ecological framework, Banyard [8] suggested taking a multilevel approach, including individual characteristics, community-level influencers, and the context of violence, when considering facilitators and barriers to bystander intervention. Bystander training programs can help potential bystanders intervene appropriately and effectively when witnessing or recognizing signs of violent incidents [11,12]. A wide variety of training methods and their effectiveness have been reported in empirical studies [13-15]. Commonly adopted training methods often include lectures, case scenario discussions, video watching, and other active and experiential learning activities such as the use of theatre [14,16]. These traditional teaching methods have demonstrated their effectiveness, such as in reducing acceptance of violence [14], increasing self-efficacy [17], and increasing the willingness to help [18].

To continue and advance the research on bystander behaviors and the effectiveness of training programs, a growing body of recent studies began to incorporate virtual reality (VR) as a tool for data collection [19-21] or as a central programming component in bystander training interventions [22]. The main advantage of VR lies in its technological advance that affords participants an immersive experiential environment. More specifically, VR surrounds individuals with a 360-degree computer-generated immersive environment, substituting real-life sensory input (predominantly visual, auditory, and tactile) using wall projections (eg, Cave Automatic Virtual Environment [CAVE]) or head-mounted display devices (eg, Oculus Rift) [23]. Besides, in a VR environment, there are one or more trackers in the room and/or attached to the user's body parts to track the user's head and body positions and movements [24]. Similar to what VR may offer, augmented reality (AR) allows creating an environment that integrates both a simulated virtual scenario and a real-world physical setting. AR superimposes digital materials over what individuals perceive in the real world [25]. In an AR environment, people can perceive and interact with virtual and physical objects, allowing for an extended real-life experience. The most common AR delivery platforms are mobile devices (eg, in Pokémon GO, one can see and play with virtual comic characters in their physical environment through the phone camera), and technology companies are experimenting with AR glasses, such as Microsoft HoloLens and Apple AR glasses. In short, VR and AR are on the reality-virtuality continuum of Milgram and Kishino [26], ranging from physical reality on one end to a completely virtual environment (ie, VR) on the other, with a blend of physical and virtual environments (ie, AR) in between [27].

The integration of VR or AR may provide new methodological and training strategies to bystander behavior research. For instance, VR provides an immersive experience that creates presence [28], shortens bystanders' psychological distance to the conflict scenarios [22], and invites bystanders to behave as if the environment is real [29]. As such, VR unites the generalizability of standardized scenarios and ecological validity by allowing bystanders to behave naturally in a controlled immersive environment [20,30]. Further, as a VR environment allows for behavioral measurements (ie, tracking), researchers can examine verbal and physical bystander behaviors rather than solely relying on self-reports, which furthers the ecological validity [20,31]. While VR simulates real-life interpersonal conflicts in a virtual environment, AR superimposes additional information on the real environment where the conflicts take place. AR has shown effectiveness in training interpersonal behaviors by digitally providing extra information about social cues [32]. In addition, mobile AR requires only a smartphone to operate and is an affordable bystander training platform for many organizations and individuals.

Aim of the Study

This review aims to comprehensively assess current evidence from empirical studies in social science and computer science databases and evaluate the use of VR or AR as a tool for studying bystander behaviors in the domain of interpersonal violence research. The reviewed results have implications and recommendations for future research in designing and implementing VR/AR technology in the area of interpersonal violence. We will examine the following questions: (1) What are the characteristics of studies using VR and/or AR in examining bystander behaviors? (2) What types of violent incidents were used in a VR and/or AR-simulated environment? (3) How are VR and/or AR used in assessing bystander behaviors or interventions? (4) What are the implications and limitations of implementing VR and/or AR as a tool in the articles?

Methods

Design

We conducted a scoping review to address the aforementioned research questions. Arksey and O'Malley's [33] framework guided the study, including "identifying the research questions; identifying relevant studies; study selection; charting the data; and collating, summarizing, and reporting the results." This review used the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping

Reviews) checklist [34]. The team consisted of members in social work, information science, data science, and user experience design.

Identifying Relevant Studies

Two team members who were librarians (JL and SCL) developed a primary search strategy in APA PsycInfo (Ovid) using a combination of text words and controlled vocabulary. The structure of the search was such that results could mention either VR or AR as long as it included bystanders (Multimedia Appendix 1). Given the long list of search terms, we only listed several examples, such as "avatars," "virtual reality," "helmet-mounted display," "Oculus Quest," "augmented reality," "bystander," "helping behaviors," and "witness." This primary search strategy was validated against an a priori test set of articles. It was peer-reviewed by another librarian, not on the team, using the Peer Review of Electronic Search Strategies (PRESS) framework. The search was then translated into the following five other databases, which were searched separately: Criminal Justice Abstracts (EBSCO), Medline (Ovid), Applied Social Sciences Index & Abstracts (ProQuest), Sociological Abstracts (ProQuest), and Scopus. All of the search queries are available in Multimedia Appendix 1. A language limit of English was applied in databases where the search retrieved over 100 results. Search results were downloaded on April 15, 2020, and uploaded to Covidence for deduplication and screening.

Study Selection

Inclusion Criteria

The team established the inclusion criteria before the screening. The team participated in training such as examining the sample search results and discussing the inclusion criteria as follows: (1) the article focused on bystanders in a violence situation; (2) the article used digital virtual simulations, such as VR and AR; and (3) the article was written in English.

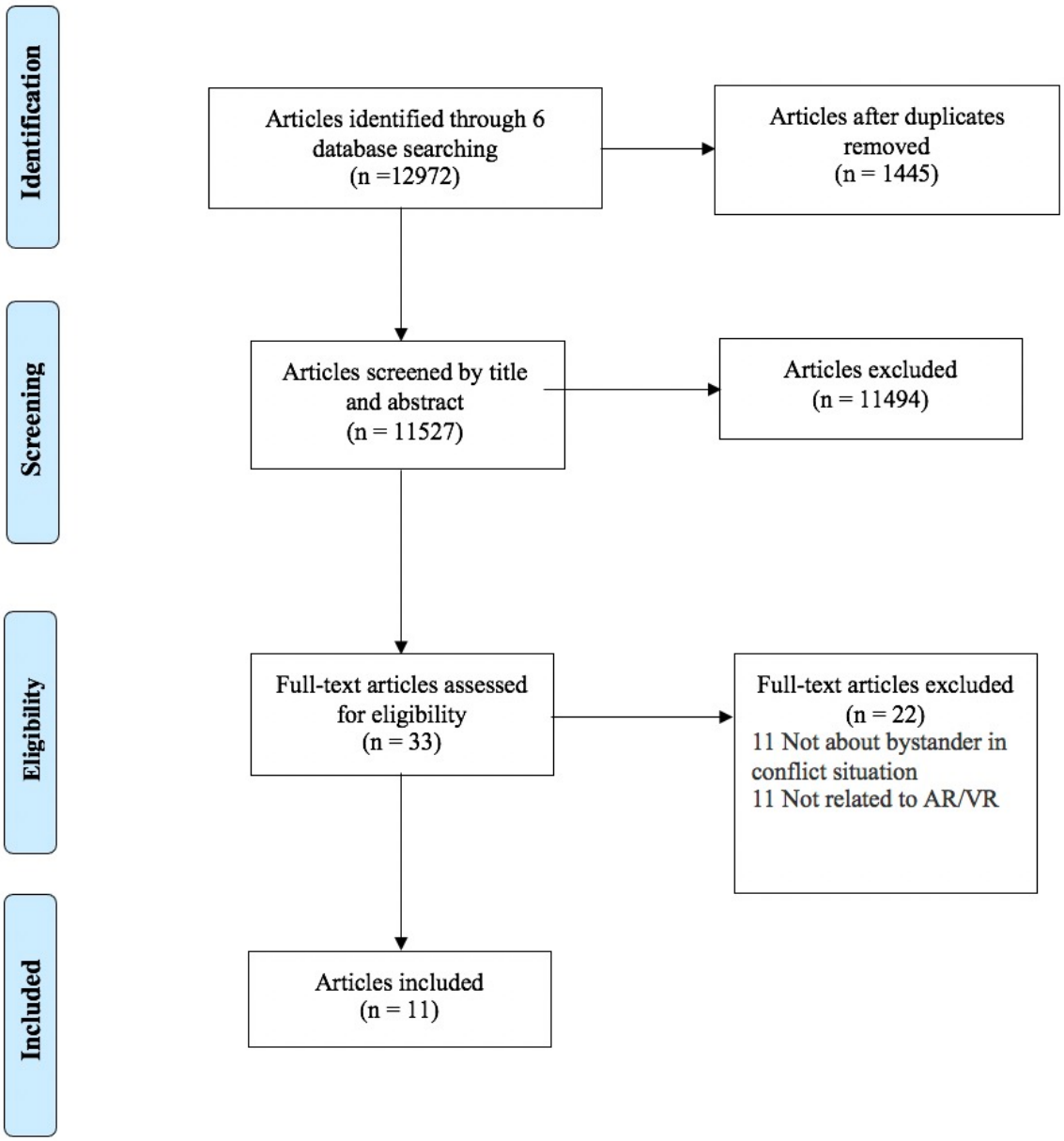
Exclusion Criteria

Articles were excluded from the review if they (1) only examined bystanders in nonconflict-related situations (eg, sudden heart attack); (2) only considered real settings, nondigital virtual settings, or non-VR/AR–related digital settings (any study only using nonimmersive virtual environments [eg, experienced through digital screens] was excluded); (3) were published in other languages; and (4) were any type of a review study, synthesis, book chapter, editorial, letter to the Editor, commentary, or opinion piece.

Selection Procedure and Search Results

As shown in Figure 1, the librarians imported 12,972 articles from six databases into Covidence. After removing 1445 duplicates, we had 11,527 articles for the title and abstract screening. First, guided by the team chair, five research assistants participated in title and abstract screening training using a random sample of 100 articles from the database. Research assistants discussed the disagreements during the training. The training was repeated twice, and we reached an interrater reliability of 100%. Second, each article was screened randomly by two out of the five research assistants. Each research assistant was assigned to review the title and abstract of 4586 articles. Third, conflict results were discussed and resolved by the team chair. We removed 11,494 ineligible articles, and a total of 33 articles remained for the full-text assessment. Fourth, each article was reviewed randomly by two out of the five research assistants. When two research assistants had any conflict, the whole team discussed the conflicts and reached a consensus. Finally, 22 articles failed to meet the inclusion criteria of the full-text screening (eg, unrelated to bystanders, unrelated to violence situations, and unrelated to VR/AR). Fifth, the team also conducted a round of hand search for key journals (eg, based on the references of the selected articles). The hand search did not add any new articles to the final data set. Therefore, our final sample for the review included 11 studies.

Figure 1. Selection procedure and search results. AR: augmented reality; VR: virtual reality.



Charting the Data

The stage of charting the data was also called "data extraction" in systematic reviews. We developed an information form to extract contents from the included articles (Tables S1, S2, and S3 in Multimedia Appendix 2), such as study location, study design, aims of the study, methodology, outcome measures, and important results. To ensure a reliable extraction process, each included article was reviewed by two authors independently, and a third author resolved any disagreements. The intercoder reliability score was excellent (Cohen κ >90% for all groups). The review process took 3 months and completed in June 2020.

Collating, Summarizing, and Reporting the Results

According to Arksey and O'Malley's [33] framework, unlike systematic reviews, our scoping review did not synthesize and evaluate the evidence quality in the studies. Instead, the scoping review summarized the research methods, sample sizes, participants, measures, geographical locations, and outcomes of the studies to report the main areas of interest and research gaps.

Results

Review Results and Summary of the Main Characteristics of the Studies

Five research assistants assessed each of the selected articles' contents and quality, and extracted and summarized the data in response to our research aim. We have charted and presented the results in Tables S1, S2, and S3 in Multimedia Appendix 2. Table S1 in Multimedia Appendix 2 presents the characteristics of the studies, including country, type of violence, participants, age, and study design. Table S2 in Multimedia Appendix 2 shows the study findings, including research questions, measurements, summary of key findings, strengths, and limitations. Table S3 in Multimedia Appendix 2 presents the design, equipment of the VR in the articles, length of the VR/AR experience, and descriptions of the scenarios.

Of the 11 studies included in the review, seven were conducted in the United States, two were conducted in the United Kingdom, one was conducted in Spain, and one did not specify the country where the research was carried out. All studies used VR as a tool to simulate certain situations of violence, and we did not find any studies that employed AR techniques. Nine studies used VR to assist observational research, and the remaining two applied VR in bystander intervention programs [22,35]. While most studies were operationalized as an experiment with participants incorporating a simulated VR environment, three studies were randomized controlled trials [20,30,35]. Different violent incident scenarios were used across these studies, including school bullying behaviors (including relational, verbal, and/or physical violence) among student peers [22,35]; sexual assault incidents that took place in dating or a potential dating relationship [20,21,35,36]; a combination of different types of assaults among friends or acquaintances [30]; and verbal and physical assaults that occurred among strangers [19,29,31,37]. In addition to the victim-aggressor relationship, we examined the relationship between victims and bystanders simulated in the VR scenarios, including school peers (n=2) [22,35], friends (n=5) [20,21,30,36,38], and strangers (n=4) [19,29,31,37]. Regarding study participants, four studies recruited middle or high school students [20,22,35,38], three studies sampled undergraduate students [21,30,36], and the four remaining studies included participants who identified as soccer fans supporting different teams [19,29,31,37].

Bullying Behaviors Among School Peers

Two studies focused on bullying behaviors among student peers in middle or high schools. They were the only two studies that included VR as a central component to deliver bullying prevention training [22,35]. Specifically, through a pseudorandomized trial, Ingram et al [22] examined whether a bullying prevention curriculum enhanced by the use of VR reduces students' bullying behaviors, including traditional bullying, cyberbullying, and relational aggression and whether the VR-enhanced intervention increases bystanders' willingness to intervene and school belonging. The study recruited 118 students from two middle schools in the United States (46 in VR-enhanced training and 72 in non-VR/regular training). To provide an immersive user experience, a set of customized VR scenarios was simulated using Google goggles [39,40]. Although the VR-enhanced condition did not reduce bullying behaviors, individuals in the VR group reported greater empathy from pretest to posttest than the non-VR group. Specifically, mediated by empathy, students in the VR condition showed decreased bullying perpetration and increased wiliness to intervene as bystanders than the control group. This study's findings suggest the promise of applying VR in school bullying prevention programs to enhance effectiveness. However, the study by McEvoy et al [35] provided somewhat different and more nuanced findings of the effectiveness of VR.

McEvoy et al [35] also conducted a randomized experimental study that aimed to compare the following three conditions: (1) a customized VR condition, (2) a noncustomized VR condition, and (3) a video condition, in order to examine whether including VR in a bullying prevention program is more effective than a less immersive video-based intervention. Seventy-eight college students participated in the study and randomly joined one of the following three conditions: (1) a 30-second video featuring a female student being verbally and physically bullied by two other female students in the school, (2) a customized VR experience based on the video in which the victim had the participants' university logo printed on her shirt, and (3) a noncustomized VR experience based on the video in which the victim had no school logo on the shirt. This study used a video from the "Be More than a Bystander campaign" by the Ad Council and the US Department of Health and Human Services [35]. The authors used Unity, a VR development software, and delivered the VR video by Oculus Rift. The results showed that participants in the video group reported higher empathy levels than those in the other two groups, and there were no differences in other measures across the three conditions. The data analysis from a follow-up focus group showed that the VR simulations needed additional interactive features to be effective. To enhance the effectiveness of incorporating VR techniques in bullying intervention programs, program developers should focus on enhancing the realism of VR-simulated environments and their autonomy in their immersive training experience [35].

Sexual Assault in Dating Relationships

Of the five studies that included sexual assault (or risks of sexual assault) in their VR-enhanced environments [20,21,35,36,38], two created sexual assault scenarios that also incorporated other types of violence [30,36]. For instance, in the study by Sargent et al [30] involving sexual assault, physical violence, stalking, and coercive controlling behavior were also simulated as part of the VR experience. In all five studies, VR was designed to assist the research rather than to develop a bystander intervention program. For instance, Sargent et al [30] designed VR simulations to determine whether VR can be used as a valid tool for assessing adolescent resistance to antisocial peer pressure. The study findings did show the psychometrical soundness of the VR simulations in research. Other studies also utilized the experiential and immersive nature of VR to collect behavioral data from participants, complementing the traditional self-report surveys. For instance, Jouriles et al [20] conducted a randomized controlled trial with 165 high school students (85 in the intervention) to assess whether the video bystander program TakeCARE increases bystander behaviors. Instead of solely relying on self-report measures to assess bystander intervening behaviors, Jouriles et al [20] also used VR simulations to collect observational data of bystander behaviors to evaluate the durability effects of TakeCARE at postintervention and 6-month follow-up assessments. Wearing goggles, all participants in the study took part in four out of nine immersive VR simulations, all of which provided them opportunities to intervene in sexual assault in a dating or potential dating relationship. Results confirmed the effectiveness of the TakeCARE bystander program to increase bystander behaviors.

Similarly, to examine whether having adverse consequences of being an active bystander was related to lower efficacy for intervening and less effective bystander behavior, Krauss et al [36] simulated a VR environment in which student participants were given opportunities to intervene in sexual and relationship violence on campus. A total of 299 first-year undergraduate

students participated in a laboratory-based assessment. Among them, 65% of the students received 20-minute bystander training (TakeCARE) given by the university before the study. In the assessment, the students participated in three 2 to 3-minute VR simulations in which they had opportunities to help the victims. Students used Oculus Rift goggles in the simulations. The results showed that negative consequences of previous bystander behaviors (being physically hurt/getting into trouble resulting from helping someone at risk of sexual assault) predicted lower bystander efficacy and effectiveness. The study also found that bystander training decreased the negative consequences of the participants. Although a limited number of different types of adverse effects were included in the analysis, the study findings indicated the importance of addressing the potential negative consequences bystanders may face or fear and educating students on how to perform safe interventions.

Verbal and Physical Assault Among Strangers

Verbal conflicts and physical assault were the focus of four studies [19,29,31,37]. All the simulated violence scenarios occurred among strangers in a public setting, such as in a bar. All these studies used VR to create an observational research environment to collect participants' behavioral data in simulated violent scenarios as bystanders. For example, Hortensius et al [29] examined whether reflexive and reflective behavioral responses to an emergency are related to later helping behavior in a violent conflict. Twenty-nine male FC Barcelona supporters (age range 18-29 years) participated in the study. They were placed in a VR-enhanced simulated conversation with the virtual victim for about 2 minutes and then witnessed a series of conflicts for 135 seconds. The technologies used in the study included the XVR programming platform [41], the virtual characters animated with HALCA software [42], and a CAVE system [43]. The results showed that the increased helping behavior was associated with a faster response to an emergency in a low cognitive load condition.

Discussion

Overview

To the best of our knowledge, our study is the first comprehensive review that evaluates studies using VR or AR as a tool for studying bystander behaviors in the domain of interpersonal violence research. This review screened 12,972 articles and assessed 11 qualified articles in full text. Evaluations of 11 eligible studies provided insights for VR/AR technology and their applications in the domain of interpersonal violence research, including VR as an observational measure, VR as an intervention tool for bystander programs, AR as a tool for the study of bystander behaviors, and the equipment and implementation of a VR system.

VR as an Observational Measure

For decades, the methodological approach for studying bystander behaviors has been participant self-reported questionnaires in response to written descriptions or videos. Our review found that studies used VR technology as a tool for observational measures (Table S1 in Multimedia Appendix 2). VR technology is a tool to present simulated scenarios. Participants are placed in immersive simulated scenarios in which their bystander behaviors can be studied. For example, participants are given opportunities to intervene in imminent violence/assault as an active bystander in these simulations (eg, girls drunk at a party, physical aggression between dating partners, and unwanted sexual activity at a party). Their bystander behaviors in the VR simulations are audio recorded, observed, and rated by coders on a 7-point scale, indicating their reactions and levels of attempting to intervene in the possible sexual or dating violence. VR provides participants an immersive virtual environment in which they genuinely interact with the avatars, which reduces the Hawthorne effect [44] (participants do not reveal their real behaviors when they realize they are being observed and studied [45]). Current evidence shows that immersive VR offers an under control environment created by computers (eg, perpetrators are made smaller in size and weaker) and, at the same time, ensures that people respond realistically. Thus, the design and implementation of these simulated environments are essential to spark any emotional, cognitive, or behavioral responses from the participants [19]. VR has opened windows of opportunities for innovative multimethods in the study of bystander behaviors in interpersonal violence situations. Further studies should continue to explore how to further improve the reliability and validity of VR as an emerging tool for observational measures.

VR as an Intervention Tool for Bystander Programs

Two studies examined whether the use of VR enhances existing bystander bullying programs [22,35]. Immersive VR provides an experimental environment that mirrors real-life situations [19]. VR-enhanced bullying scenarios are designed for participants (adolescents) to simulate real-life bullying situations. For example, the participant wears an Oculus Rift RV headset and headphones and experiences the designed scenarios (witnessing a female student being verbally and physically bullied by other students in the school hallway between classes) [35] (Table S3 in Multimedia Appendix 2). The contents of scenarios are designed to resemble bullying prevention program videos (eg, audio and message screen) or be informed by the empirical literature [22]. The benefits of VR over video curriculum as a tool for violence prevention are suggested. Unlike the video curriculum, participants can look around in immersive VR simulations, even though they cannot interact with the avatars. In addition to the video curriculum, the VR-enhanced prevention tool allows participants to experience different perspectives in the simulation, such as being bystanders, being victims, and being adults to intervene. VR is an ecologically valid environment in which researchers can overcome ethical issues in violence studies and prevent potential real physical danger to participants [19].

AR as a Tool for Bystander Behaviors

Our review did not find any empirical studies that assessed the use of AR for bystander behaviors in interpersonal violence situations. AR enables integrating physical and virtual elements into one view by allowing participants to merge the virtual component into the real physical world. AR is different from VR, and it enables participants to be in an immersive virtual world [46]. Previous research in AR indicates that challenges

exist in the social acceptance of AR, such as privacy concerns [47,48]. For example, the subtle design of the Google glass makes bystanders around feel privacy destruction and intrusion [49,50]. This review suggests that future studies may examine the feasibility and effectiveness of using AR technology to study bystander behaviors in the domain of interpersonal violence research.

Equipment and Implementation of the VR System

The results of the reviewed articles suggest that future studies should give attention to what technology is being used as an experimental tool and an observational measure, and how VR platforms are set up because these could affect the results. We have identified several key features of the design and equipment for using VR to study bystander behaviors in interpersonal violence research. These include designed virtual avatars and VR scenarios, a CAVE-like system (automatic virtual environment), glasses for immersive 3D vision, including headphones and a microphone, a head tracker, a video camera for recording, and a programming platform (or simulations created and coded by the authors).

In our review, only four studies [19,29,31,37] used a CAVE-like system. For example, Rovira et al [19] used Trimension ReacTor, which has three back-projected acrylic screens (front, left, and right walls; 3 m × 2.2 m) and a floor screen projected from a ceiling-mounted projector (3 m × 3 m) [51,52]. The CAVE system used in these studies has variations in pixel resolution and Hertz (monitor's refresh rate, 60 Hz means the monitor refreshes its image 60 times per second), which have impacts on participants' intervention behaviors [37].

Without employing a sophisticated CAVE-like system, head-mounted display devices can be used only to design a VR-based project for studying and measuring bystander behaviors. Different glasses are used to synchronize with the projectors in the VR system, such as Oculus Rift [30,35,36], Crystal Eyes shutter glasses [19,29,31,37], and Daydream goggles [22]. Three studies did not indicate the brand of the goggles used [20,21,38].

Limitations

There are several limitations in this review. First, we only included articles using VR and/or AR as a tool for studying bystander behaviors and therefore excluded studies that examined the engagement of violence on these platforms, such as harassment in social VR [53] and violence in video games [54]. Second, there were variations in the design and equipment of VR in these articles, making it challenging to perform cross-study comparisons. Despite these limitations, our study is the first to review the use of VR as a tool for bystander behaviors in interpersonal violence research.

Future Research: Recommendations for the Use of VR in Interpersonal Violence Research

Further studies are needed to provide a rich understanding of the use of VR in the domain of interpersonal violence research. First, future studies should further compare the effectiveness of VR as an intervention tool with the well-established bystander intervention curriculum based on videos. Second, the cost of VR may prevent its widespread implementation. Future studies should further explore potential low-cost VR designs and equipment. Third, future research may consider collaborating with human-computer interaction experts to examine whether technology's quality is associated with efficacy. For example, although current evidence cannot quantify the effects of display resolution and luminance on people's responses, detailed facial expressions of the avatars (eg, 1400 × 1050 pixels, 100-Hz refresh rate, 3150 lumens, and digital light processing projectors) encourage more empathic bystander responses [37]. Such recommendations can help determine whether the development and implementation of VR technology are cost-effective in research to increase bystanders' intent to intervene in violent situations. Fourth, future research should explore the variables that mediate the use of VR, such as psychological distance in bullying [22], photorealistic graphics [35], participants' familiarity with VR technology, and variations across administrations [21]. Fifth, given the small sample size in all reviewed articles, no evidence was provided with implementation in adults and older adults. Future studies may consider expanding the age range to a broader population to increase the generalizability of the findings.

Conclusion

There remain considerable gaps in the literature regarding the use of VR technology, notably AR, as a tool for studying bystander behaviors in the domain of interpersonal violence research. The current evidence suggests the effectiveness of VR as an observational measure in addition to self-reported questionnaires and as an intervention tool compared with video-only bystander programs. A limited number of studies exist, and it justifies further research efforts in this area.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete search strategy.
[DOCX File , 20 KB - jmir_v23i2e25322_app1.docx]

Multimedia Appendix 2
Summaries of the studies, research findings, and virtual reality/augmented reality designs.
[DOCX File , 39 KB - jmir_v23i2e25322_app2.docx]

References

1. Global status report on violence prevention 2014. World Health Organization. URL: https://www.who.int/violence_injury_prevention/violence/status_report/2014/en/ [accessed 2020-07-01]
2. Krug E, Mercy J, Dahlberg L, Zwi A. The world report on violence and health. The Lancet 2002 Oct;360(9339):1083-1088 [FREE Full text] [doi: 10.1016/s0140-6736(02)11133-0]
3. Carey K, Norris A, Durney S, Shepardson R, Carey M. Mental health consequences of sexual assault among first-year college women. J Am Coll Health 2018;66(6):480-486 [FREE Full text] [doi: 10.1080/07448481.2018.1431915] [Medline: 29405862]
4. Peter-Hagene L, Ullman S. Longitudinal Effects of Sexual Assault Victims' Drinking and Self-Blame on Posttraumatic Stress Disorder. J Interpers Violence 2016 Mar 07:83-93 [FREE Full text] [doi: 10.1177/0886260516636394] [Medline: 26956436]
5. Trevillion K, Oram S, Feder G, Howard L. Experiences of domestic violence and mental disorders: a systematic review and meta-analysis. PLoS One 2012;7(12):e51740 [FREE Full text] [doi: 10.1371/journal.pone.0051740] [Medline: 23300562]
6. Bennett S, Banyard V, Garnhart L. To act or not to act, that is the question? Barriers and facilitators of bystander intervention. J Interpers Violence 2014 Feb;29(3):476-496 [FREE Full text] [doi: 10.1177/0886260513505210] [Medline: 24097909]
7. Padgett S, Notar CE. Bystanders are the Key to Stopping Bullying. Universal Journal of Educational Research 2013;1(2):33-41 [FREE Full text] [doi: 10.13189/ujer.2013.010201]
8. Banyard V. Who will help prevent sexual violence: Creating an ecological model of bystander intervention. Psychology of Violence 2011;1(3):216-229 [FREE Full text] [doi: 10.1037/a0023739]
9. McMahon S, Peterson N, Winter S, Palmer J, Postmus J, Koenick R. Predicting Bystander Behavior to Prevent Sexual Assault on College Campuses: The Role of Self-Efficacy and Intent. Am J Community Psychol 2015 Sep;56(1-2):46-56 [FREE Full text] [doi: 10.1007/s10464-015-9740-0] [Medline: 26194588]
10. Kania R, Cale J. Preventing Sexual Violence Through Bystander Intervention: Attitudes, Behaviors, Missed Opportunities, and Barriers to Intervention Among Australian University Students. J Interpers Violence 2018 Mar 01:886260518764395 [FREE Full text] [doi: 10.1177/0886260518764395] [Medline: 29557711]
11. Jouriles E, Krauss A, Vu N, Banyard V, McDonald R. Bystander programs addressing sexual violence on college campuses: A systematic review and meta-analysis of program outcomes and delivery methods. J Am Coll Health 2018;66(6):457-466 [FREE Full text] [doi: 10.1080/07448481.2018.1431906] [Medline: 29405865]
12. Storer H, Casey E, Herrenkohl T. Efficacy of Bystander Programs to Prevent Dating Abuse Among Youth and Young Adults: A Review of the Literature. Trauma Violence Abuse 2016 Jul;17(3):256-269 [FREE Full text] [doi: 10.1177/1524838015584361] [Medline: 25951840]
13. Alegría-Flores K, Raker K, Pleasants R, Weaver M, Weinberger M. Preventing Interpersonal Violence on College Campuses: The Effect of One Act Training on Bystander Intervention. J Interpers Violence 2017 Apr;32(7):1103-1126 [FREE Full text] [doi: 10.1177/0886260515587666] [Medline: 26002875]
14. Coker A, Bush H, Brancato C, Clear E, Recktenwald E. Bystander Program Effectiveness to Reduce Violence Acceptance: RCT in High Schools. J Fam Violence 2019;34(3):153-164 [FREE Full text] [doi: 10.1007/s10896-018-9961-8] [Medline: 30956390]
15. Midgett A, Doumas D, Sears D, Lundquist A, Hausheer R. A Bystander Bullying Psychoeducation Program With Middle School Students: A Preliminary Report. TPC 2015 Dec;5(4):486-500 [FREE Full text] [doi: 10.15241/am.5.4.486]
16. Mujal G, Taylor M, Fry J, Gochez-Kerr T, Weaver N. A Systematic Review of Bystander Interventions for the Prevention of Sexual Violence. Trauma Violence Abuse 2019 Jun 16:1524838019849587 [FREE Full text] [doi: 10.1177/1524838019849587] [Medline: 31204606]
17. Katz J, Moore J. Bystander education training for campus sexual assault prevention: an initial meta-analysis. Violence Vict 2013;28(6):1054-1067 [FREE Full text] [doi: 10.1891/0886-6708.vv-d-12-00113] [Medline: 24547680]
18. Amar A, Sutherland M, Kesler E. Evaluation of a bystander education program. Issues Ment Health Nurs 2012 Dec;33(12):851-857 [FREE Full text] [doi: 10.3109/01612840.2012.709915] [Medline: 23215986]
19. Rovira A, Swapp D, Spanlang B, Slater M. The Use of Virtual Reality in the Study of People's Responses to Violent Incidents. Front Behav Neurosci 2009;3:59 [FREE Full text] [doi: 10.3389/neuro.08.059.2009] [Medline: 20076762]
20. Jouriles E, McDonald R, Rosenfield D, Sargent K. Increasing bystander behavior to prevent adolescent relationship violence: A randomized controlled trial. J Consult Clin Psychol 2019 Jan;87(1):3-15 [FREE Full text] [doi: 10.1037/ccp0000355] [Medline: 30474990]
21. Jouriles E, Kleinsasser A, Rosenfield D, McDonald R. Measuring bystander behavior to prevent sexual violence: Moving beyond self reports. Psychology of Violence 2016 Jan;6(1):73-81 [FREE Full text] [doi: 10.1037/a0038230]
22. Ingram K, Espelage D, Merrin G, Valido A, Heinhorst J, Joyce M. Evaluation of a virtual reality enhanced bullying prevention curriculum pilot trial. J Adolesc 2019 Feb;71:72-83 [FREE Full text] [doi: 10.1016/j.adolescence.2018.12.006] [Medline: 30639666]
23. Guttentag D. Virtual reality: Applications and implications for tourism. Tourism Management 2010 Oct;31(5):637-651 [FREE Full text] [doi: 10.1016/j.tourman.2009.07.003]

24. Riva G, Baños RM, Botella C, Mantovani F, Gaggioli A. Transforming Experience: The Potential of Augmented Reality and Virtual Reality for Enhancing Personal and Clinical Change. Front Psychiatry 2016;7:164 [FREE Full text] [doi: 10.3389/fpsyt.2016.00164] [Medline: 27746747]
25. Azuma R. A Survey of Augmented Reality. Presence: Teleoperators & Virtual Environments 1997 Aug;6(4):355-385 [FREE Full text] [doi: 10.1162/pres.1997.6.4.355]
26. Milgram P, Kishino F. A Taxonomy of Mixed Reality Visual Displays. IEICE Trans Inf & Syst 1994;E77-D:1321-1329.
27. Flavián C, Ibáñez-Sánchez S, Orús C. The impact of virtual, augmented and mixed reality technologies on the customer experience. Journal of Business Research 2019 Jul;100:547-560 [FREE Full text] [doi: 10.1016/j.jbusres.2018.10.050]
28. Cipresso P, Giglioli I, Raya M, Riva G. The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature. Front Psychol 2018;9:2086 [FREE Full text] [doi: 10.3389/fpsyg.2018.02086] [Medline: 30459681]
29. Hortensius R, Neyret S, Slater M, de Gelder B. The relation between bystanders' behavioral reactivity to distress and later helping behavior during a violent conflict in virtual reality. PLoS One 2018;13(4):e0196074 [FREE Full text] [doi: 10.1371/journal.pone.0196074] [Medline: 29672638]
30. Sargent K, Jouriles E, Chmielewski M, McDonald R. Using Virtual Reality to Create an Observational Assessment of Adolescent Resistance to Antisocial Peer Pressure. J Clin Child Adolesc Psychol 2020 Aug 24;49(2):178-189 [FREE Full text] [doi: 10.1080/15374416.2018.1504296] [Medline: 30142280]
31. Slater M, Rovira A, Southern R, Swapp D, Zhang JJ, Campbell C, et al. Bystander responses to a violent incident in an immersive virtual environment. PLoS One 2013;8(1):e52766 [FREE Full text] [doi: 10.1371/journal.pone.0052766] [Medline: 23300991]
32. Chen C, Lee I, Lin L. Augmented reality-based video-modeling storybook of nonverbal facial cues for children with autism spectrum disorder to improve their perceptions and judgments of facial expressions and emotions. Computers in Human Behavior 2016 Feb;55:477-485 [FREE Full text] [doi: 10.1016/j.chb.2015.09.033]
33. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
34. Tricco A, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
35. McEvoy K, Oyekoya O, Ivory A, Ivory J. Through the eyes of a bystander: The promise and challenges of VR as a bullying prevention tool. 2016 Presented at: 2016 IEEE Virtual Reality (VR); July 07, 2016; Greenville, SC, USA. [doi: 10.1109/vr.2016.7504737]
36. Krauss A, Jouriles E, Yule K, Grych J, Sargent K, Banyard V. Adverse Consequences to Assisting Victims of Campus Violence: Initial Investigations Among College Students. J Interpers Violence 2017 Dec 01:886260517746944 [FREE Full text] [doi: 10.1177/0886260517746944] [Medline: 29295036]
37. Rovira A, Swapp D, Southern R, Zhang J, Slater M. The impact of enhanced projector display on the responses of people to a violent scenario in immersive virtual reality. 2013 Presented at: 2013 IEEE Virtual Reality (VR); March 18-20, 2013; Lake Buena Vista, FL, USA. [doi: 10.1109/vr.2013.6549350]
38. Jouriles E, Rosenfield D, Yule K, Sargent K, McDonald R. Predicting High-School Students' Bystander Behavior in Simulated Dating Violence Situations. J Adolesc Health 2016 Mar;58(3):345-351 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.11.009] [Medline: 26794432]
39. Daydream. URL: https://arvr.google.com/intl/en_ca/daydream/ [accessed 2020-08-01]
40. Dascalu M, Sergean B, Ferche OM, Moldoveanu A. Experiential Learning VR System for Studying Computer Architecture. Romanian Journal of Human-Computer Interaction 2017;10:197-215.
41. Tecchia F. A Flexible Framework for Wide‐Spectrum VR Development. Presence: Teleoperators and Virtual Environments 2010 Aug 01;19(4):302-312. [doi: 10.1162/pres_a_00002]
42. Gillies M, Spanlang B. Comparing and Evaluating Real Time Character Engines for Virtual Environments. Presence: Teleoperators and Virtual Environments 2010 Apr 01;19(2):95-117. [doi: 10.1162/pres.19.2.95]
43. Cruz-Neira C, Sandin D, DeFanti T, Kenyon R, Hart J. The CAVE: audio visual experience automatic virtual environment. Commun. ACM 1992 Jun;35(6):64-72 [FREE Full text] [doi: 10.1145/129888.129892]
44. Bombari D, Schmid Mast M, Canadas E, Bachmann M. Studying social interactions through immersive virtual environment technology: virtues, pitfalls, and future challenges. Front Psychol 2015;6:869 [FREE Full text] [doi: 10.3389/fpsyg.2015.00869] [Medline: 26157414]
45. McCambridge J, Witton J, Elbourne D. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]
46. Baus O, Bouchard S. Moving from virtual reality exposure-based therapy to augmented reality exposure-based therapy: a review. Front Hum Neurosci 2014;8:112 [FREE Full text] [doi: 10.3389/fnhum.2014.00112] [Medline: 24624073]
47. Denning T, Dehlawi Z, Kohno T. In situ with bystanders of augmented reality glasses: perspectives on recording and privacy-mediating technologies. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. In;

2014 Presented at: CHI '14: The SIGCHI Conference on Human Factors in Computing Systems; April 2014; Toronto, Ontario, Canada p. 2377-2386. [doi: 10.1145/2556288.2557352]
48. Van Krevelen D, Poelman R. A Survey of Augmented Reality Technologies, Applications and Limitations. IJVR 2010 Jan 01;9(2):1-20 [FREE Full text] [doi: 10.20870/ijvr.2010.9.2.2767]
49. John Naughton. The Guardian. URL: https://www.theguardian.com/profile/johnnaughton [accessed 2021-01-29]
50. Singhal S, Neustaedter C, Schiphorst T, Tang A, Patra A, Pan R. You are Being Watched: Bystanders' Perspective on the Use of Camera Devices in Public Spaces. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI EA '16: The 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems; May 2016; San Jose, CA, USA p. 3197-3203. [doi: 10.1145/2851581.2892522]
51. Pan X, Gillies M, Slater M. Virtual Character Personality Influences Participant Attitudes and Behavior â€“ An Interview with a Virtual Human Character about Her Social Anxiety. Front. Robot. AI 2015 Feb 10;2 [FREE Full text] [doi: 10.3389/frobt.2015.00001]
52. Slater M, Antley A, Davison A, Swapp D, Guger C, Barker C, et al. A virtual reprise of the Stanley Milgram obedience experiments. PLoS One 2006 Dec 20;1:e39 [FREE Full text] [doi: 10.1371/journal.pone.0000039] [Medline: 17183667]
53. Shriram K, Schwartz R. All are welcome: Using VR ethnography to explore harassment behavior in immersive social virtual reality. 2017 Presented at: 2017 IEEE Virtual Reality (VR); March 18-22, 2017; Los Angeles, CA, USA. [doi: 10.1109/vr.2017.7892258]
54. Fox J, Tang W. Women's experiences with general and sexual harassment in online video games: Rumination, organizational responsiveness, withdrawal, and coping strategies. New Media & Society 2016 Mar 08;19(8):1290-1307 [FREE Full text] [doi: 10.1177/1461444816635778]

Abbreviations

AR: augmented reality
CAVE: Cave Automatic Virtual Environment
VR: virtual reality

Edited by G Eysenbach; submitted 27.10.20; peer-reviewed by C Emezue, H Söbke; comments to author 17.11.20; revised version received 30.12.20; accepted 20.01.21; published 15.02.21.

Please cite as:
Xue J, Hu R, Zhang W, Zhao Y, Zhang B, Liu N, Li SC, Logan J
Virtual Reality or Augmented Reality as a Tool for Studying Bystander Behaviors in Interpersonal Violence: Scoping Review
J Med Internet Res 2021;23(2):e25322
URL: http://www.jmir.org/2021/2/e25322/
doi:10.2196/25322
PMID:33587044

Original Paper

An 8-Week Self-Administered At-Home Behavioral Skills-Based Virtual Reality Program for Chronic Low Back Pain: Double-Blind, Randomized, Placebo-Controlled Trial Conducted During COVID-19

Laura M Garcia[1], PhD; Brandon J Birckhead[1], MD, MHDS; Parthasarathy Krishnamurthy[2], MBA, PhD; Josh Sackman[1], MBA; Ian G Mackey[1], BA; Robert G Louis[3], MD; Vafi Salmasi[4], MD, MSc; Todd Maddox[1], PhD; Beth D Darnall[4], PhD

[1]AppliedVR, Inc, Los Angeles, CA, United States

[2]CT Bauer College of Business, University of Houston, Houston, TX, United States

[3]Division of Neurosurgery, Pickup Family Neurosciences Institute, Hoag Memorial Hospital, Newport Beach, CA, United States

[4]Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Palo Alto, CA, United States

Corresponding Author:
Beth D Darnall, PhD
Department of Anesthesiology, Perioperative and Pain Medicine
Stanford University School of Medicine
1070 Arastradero Road
Ste 200, MC5596
Palo Alto, CA, 94304
United States
Phone: 1 15035778377
Email: bdarnall@stanford.edu

Abstract

Background: Chronic low back pain is the most prevalent chronic pain condition worldwide and access to behavioral pain treatment is limited. Virtual reality (VR) is an immersive technology that may provide effective behavioral therapeutics for chronic pain.

Objective: We aimed to conduct a double-blind, parallel-arm, single-cohort, remote, randomized placebo-controlled trial for a self-administered behavioral skills-based VR program in community-based individuals with self-reported chronic low back pain during the COVID-19 pandemic.

Methods: A national online convenience sample of individuals with self-reported nonmalignant low back pain with duration of 6 months or more and with average pain intensity of 4 or more/10 was enrolled and randomized 1:1 to 1 of 2 daily (56-day) VR programs: (1) EaseVRx (immersive pain relief skills VR program); or (2) Sham VR (2D nature content delivered in a VR headset). Objective device use data and self-reported data were collected. The primary outcomes were the between-group effect of EaseVRx versus Sham VR across time points, and the between–within interaction effect representing the change in average pain intensity and pain-related interference with activity, stress, mood, and sleep over time (baseline to end-of-treatment at day 56). Secondary outcomes were global impression of change and change in physical function, sleep disturbance, pain self-efficacy, pain catastrophizing, pain acceptance, pain medication use, and user satisfaction. Analytic methods included intention-to-treat and a mixed-model framework.

Results: The study sample was 179 adults (female: 76.5%, 137/179; Caucasian: 90.5%, 162/179; at least some college education: 91.1%, 163/179; mean age: 51.5 years [SD 13.1]; average pain intensity: 5/10 [SD 1.2]; back pain duration ≥5 years: 67%, 120/179). No group differences were found for any baseline variable or treatment engagement. User satisfaction ratings were higher for EaseVRx versus Sham VR (P<.001). For the between-groups factor, EaseVRx was superior to Sham VR for all primary outcomes (highest P value=.009), and between-groups Cohen d effect sizes ranged from 0.40 to 0.49, indicating superiority was moderately clinically meaningful. For EaseVRx, large pre–post effect sizes ranged from 1.17 to 1.3 and met moderate to substantial clinical importance for reduced pain intensity and pain-related interference with activity, mood, and stress. Between-group comparisons for Physical Function and Sleep Disturbance showed superiority for the EaseVRx group versus the Sham VR group

(*P*=.022 and .013, respectively). Pain catastrophizing, pain self-efficacy, pain acceptance, prescription opioid use (morphine milligram equivalent) did not reach statistical significance for either group. Use of over-the-counter analgesic use was reduced for EaseVRx (*P*<.01) but not for Sham VR.

Conclusions: EaseVRx had high user satisfaction and superior and clinically meaningful symptom reduction for average pain intensity and pain-related interference with activity, mood, and stress compared to sham VR. Additional research is needed to determine durability of treatment effects and to characterize mechanisms of treatment effects. Home-based VR may expand access to effective and on-demand nonpharmacologic treatment for chronic low back pain.

Trial Registration: ClinicalTrials.gov NCT04415177; https://clinicaltrials.gov/ct2/show/NCT04415177

International Registered Report Identifier (IRRID): RR2-10.2196/25291

(J Med Internet Res 2021;23(2):e26292) doi:10.2196/26292

KEYWORDS

virtual reality; low back pain; opioids; chronic pain; behavioral health; pain treatment, randomized controlled trial; COVID-19

Introduction

Chronic low back pain (cLBP) is the most prevalent chronic pain condition worldwide [1]. cLBP can be disabling, costly, and confer suffering to individuals and their families. The incidence and prevalence of cLBP continue to rise despite increasing use of medical treatments such as pharmacology and surgical procedures [2].

An expert evidence review and consensus panel recommended pain education and cognitive behavioral therapy (CBT) as first-line treatments for cLBP [3] with both modalities supplying self-help information for back pain. Beyond the context of back pain, the Centers for Disease Control and Prevention [4] (CDC) and the Centers for Medicare & Medicaid Services (CMS) [5] have recommended nonpharmacologic therapies as first-line treatments for chronic pain, with the latter citing a need to improve patient access to effective treatment options.

CBT for chronic pain engages participants in active pain and symptom self-management [4,6,7]. In group settings, CBT is delivered by trained therapists and typically involves 8-12 two-hour treatment sessions (16-24 hours total treatment time). Manualized session content includes health and pain education; skills training in goal setting, problem solving and action planning; self-regulatory techniques (eg, relaxation, mindfulness); cognitive techniques (eg, thought monitoring and restructuring unhelpful thoughts); and functional goal setting. While CBT has not shown efficacy for reducing pain intensity, it has small to moderate effects for reducing depressive symptoms [7], pain bothersomeness [6,7], and pain catastrophizing [6,7] (Darnall et al, unpublished data) in mixed etiology chronic pain as well as cLBP. Despite demonstrated efficacy for these multisession behavioral pain treatments, access to care remains poor due to barriers such as few trained and available local therapists, health insurance limits, and burdens associated with travel and treatment time [8]. Because of the scope and impact of cLBP, there is an urgent need for effective, accessible, low-risk treatments that are acceptable to people who have back pain. Improved access to behavioral pain care is particularly salient within the context of reduced opioid prescribing for chronic pain nationally [9].

On-demand digital therapeutics may provide home-based access to pain education and skills-based pain self-management. Particularly during the COVID-19 pandemic, home-based behavioral pain care has gained interest, importance, and engagement among patients [10]. Home-based digital pain treatment options include those involving therapist instruction [11,12] (Ziadni et al, unpublished data), as well as fully automated behavioral programs. For the latter, the portfolio of self-treatment options includes computer applications for symptom tracking, education, and treatment [13]; web-based programs that include self-paced multisession skills-based pain management learning modules [14]; and virtual reality (VR) immersive treatment [15].

VR treatment involves using headset devices that fully restrict the vision field to content displayed inside the headset screen; auditory perception is not fully restricted, though the corresponding device-delivered auditory content commands attention. As a treatment modality, VR provides a unique environment comprising 3D visually immersive experiences that are enriched with stereo sounds and elements such as rich colors and scenic environments that enhance elicitation of desired states of arousal and affect. Within the therapeutic context, VR may be flexibly designed and tailored to address the needs of specific conditions (eg, anxiety, depression, pain) [15-19]. The multisensory immersive VR environment stimulates the visual, auditory, and proprioceptive senses, thus engendering the perception that the user is physically located within the virtual environment they are viewing in the headset [20,21]. Mechanistically, the integrated, multisensory, and immersive properties of VR are thought to enhance treatment effects. For example, results from phobia treatment research has suggested noninferiority of VR treatment compared to treatment with a live therapist [18].

In terms of pain, evidence from multiple independent research groups suggests that VR is effective for managing acute pain [22], including pain evoked during medical procedures [23-28], burn wound care [29,30], and in hospitalized patients [31,32]. Researchers of a randomized controlled trial (RCT) conducted in hospitalized patients found that VR yielded the highest efficacy in patients reporting the most severe pain (≥7/10), thereby underscoring its potent analgesic potential [31].

The scientific literature on VR for chronic pain includes studies conducted in complex regional pain syndrome [33], chronic headache/migraine [34], fibromyalgia [35,36], and chronic

musculoskeletal pain [37,38]. Two recent reviews and meta-analyses reported VR efficacy for reducing pain and disability (physical rehabilitation) for painful spinal conditions [39] and for orthopedic rehabilitation [40]. Such rehabilitation studies may apply interactive VR in isolation or with kinematic training [41,42]. In addition to small sample sizes, the literature for VR in chronic pain remains limited by studies conducted in experimental or clinical settings (versus home-based and pragmatic studies), a lack of placebo-controlled studies, and studies that have yielded low-quality evidence [39]. Additionally, the literature has been largely restricted to VR content involving distraction or physical rehabilitation and kinematic exercises with little or no content on active behavioral pain management skills acquisition. To address several of these evidence gaps, our group recently conducted an RCT of a 21-day VR program that included chronic pain education and pain relief skills such as diaphragmatic breathing and relaxation response training, and cognition and emotion regulation techniques [15]. Individuals with cLBP or fibromyalgia were randomized to receive either the 21-day VR treatment program or the same treatment content delivered in audio-only format (N=74). At posttreatment, the VR skills-based treatment group evidenced superior reductions in pain intensity and pain-related interference with activity, sleep, mood, and stress compared to the audio treatment group, with results strengthening after 2 weeks. Similar treatment engagement rates between treatment groups supported a conclusion that the immersive effects of VR yielded superior outcomes [15].

This study builds on this prior work and extends it in several ways. First, the VR treatment program being tested (EaseVRx) is 56 days in length, thereby aligning more with traditional and reimbursable behavioral medicine programs such as 8-week chronic pain CBT or mindfulness programs. Second, the VR content was enriched with interoceptive entrainment techniques designed to enhance biofeedback response and learning. Third, the therapeutic VR program includes expanded pain neuroscience education, as well as principles and elements drawn from CBT, mindfulness, and acceptance-based treatments for chronic pain. Fourth, the study includes a VR sham comparator to control for the novelty of the technology and placebo effects. Fifth, data for analgesic medication use were collected.

Our objective was to conduct a placebo-controlled RCT in community-based individuals with cLBP assigned to receive one of two 56-day treatment programs: therapeutic VR (EaseVRx) or Sham VR [43]. We hypothesized that participants assigned to therapeutic VR would evidence superior outcomes for all baseline to posttreatment comparisons compared to participants assigned to Sham VR.

Methods

Study Protocol

The study protocol is published elsewhere and provides additional detail [43]. We conducted a single-cohort, double-blinded (participant and analysts), placebo-controlled randomized clinical trial in an online convenience national sample of community-based individuals with self-reported cLBP. Study participants were participating in a longer study of 8.5 months' duration that involves multiple additional posttreatment assessments not reported here. The study was approved by the Western Institutional Review Board on July 2, 2020, and data collection for this report was completed in November 2020. This report is constrained to the end of treatment time point (day 56).

Study Sample, Setting, Recruitment, and Participant Compensation

Community-based individuals with cLBP were recruited nationally through chronic pain organizations (eg, American Chronic Pain Association) and through Facebook online advertisements. Additionally, study advertisements were emailed to professional contacts at several medical clinics with requests to forward among medical colleagues nationally. All study advertisements directed interested individuals with cLBP to the study website for information and invitation to complete an automated online eligibility form that screened for inclusion and exclusion criteria (Textbox 1). This screening process was completed over the phone for 1 individual due to technical difficulties. Individuals determined to be eligible for the study were invited to participate in a study examining the effectiveness of an 8-week VR wellness program in helping them manage chronic lower back pain. If willing to participate, participants completed an electronic informed consent (eConsent; see Multimedia Appendix 1 for the study consent form) and provided their e-signature to complete their enrollment in the study. Figure 1 displays the participant study activities.

Enrolled participants completed a baseline survey and then received 3 pain surveys over a 2-week pretreatment assessment period. Those who completed at least one survey during this pretreatment period progressed to the treatment phase of the study, which included an 8-week VR program (therapeutic VR or Sham VR), twice-weekly surveys during the VR treatment phase, and a final survey administered at treatment completion on day 56. All study procedures occurred remotely.

Study participation was compensated in 2 ways. First, participants received US $6 per completed survey (US $126 possible; prorated) in the form of an Amazon eGift Card after the day 56 survey and upon return of their VR headset (prepaid shipping containers were provided). Second, all participants were eligible to receive a gift VR headset 6 months after their completion of treatment if they completed 16 or more study surveys, confirmed their interest in receiving a VR headset, and returned their VR headset. A total of 77 study participants met these criteria and will receive headsets (Textbox 1).

Textbox 1. Study inclusion and exclusion criteria.

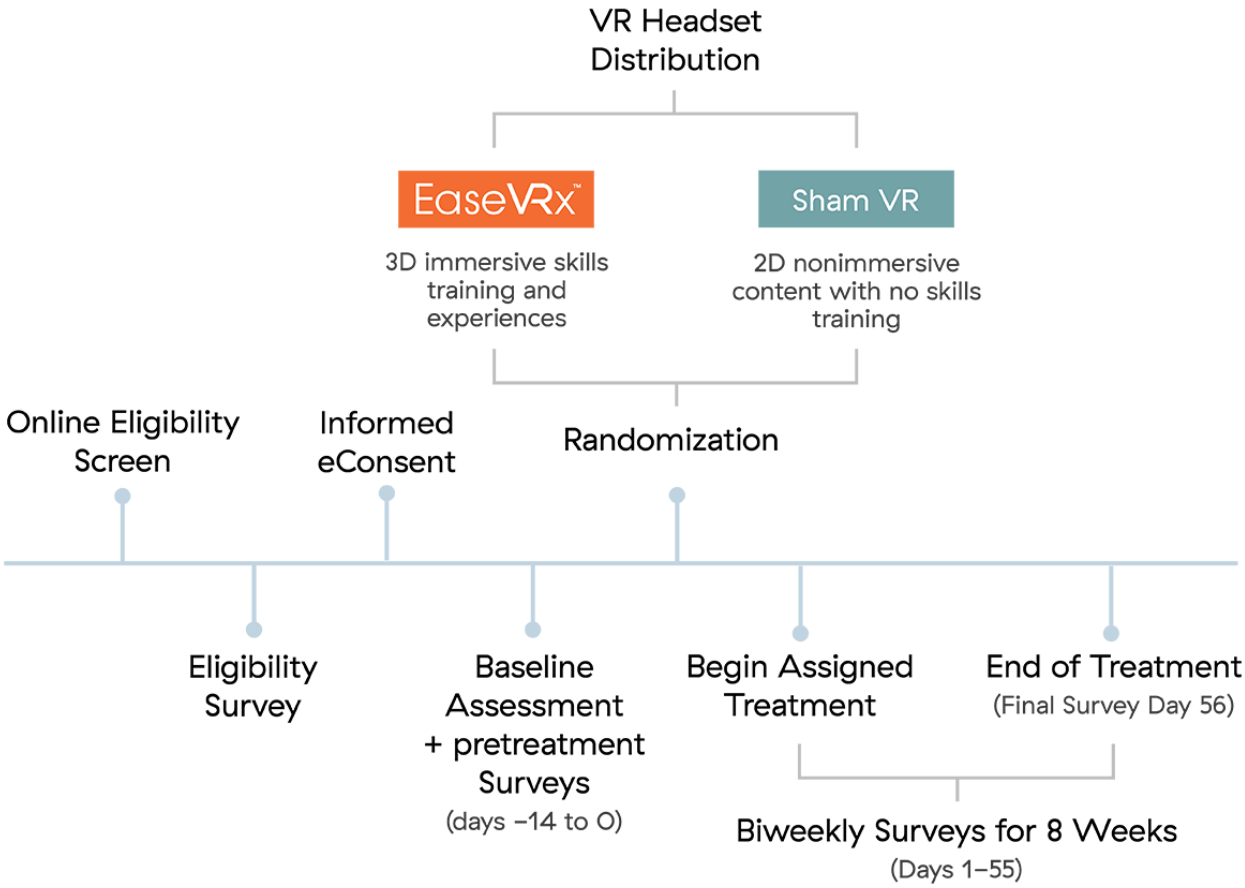
Inclusion criteria

1. Men and women aged 18-85.
2. Self-reported diagnosis of chronic low back pain without radicular symptoms.
3. Chronic low back pain duration of 6 months or more.
4. Average pain intensity of 4 or more out of 10 for the past month.
5. English fluency.
6. Willing to comply with study procedures/restrictions.
7. Access to Wi-Fi.
8. Implicit de facto internet and computer literacy.

Exclusion criteria

1. Gross cognitive impairment.
2. Current or prior diagnosis of epilepsy, seizure disorder, dementia, migraines, or other neurological diseases that may prevent the use of virtual reality or adverse effects.
3. Medical condition predisposing to nausea or dizziness.
4. Hypersensitivity to flashing light or motion.
5. No stereoscopic vision or severe hearing impairment.
6. Injury to eyes, face, or neck that impedes comfortable use of virtual reality.
7. Cancer-related pain.
8. Moderate depressive symptoms as indicated by the Patient Health Questionnaire-2 (PHQ-2 [44,45]) depression screen score of 2 or more.
9. Previous use of EaseVRx for pain.
10. Current or recent completion of participation (past 2 months) in any interventional research study.
11. Currently pregnant or planning to become pregnant during the study period.
12. Not expected to have access to Wi-Fi during the study period.
13. Currently works at or has an immediate family member who works for a digital health company or pharmaceutical company that provides treatments for acute or chronic pain.

Figure 1. Participant activities.



Randomization and Blinding

Participants were randomized 1:1 and allocated to 1 of 2 treatment groups: (1) the 56-day pain relief skills VR program (EaseVRx) or (2) a 56-day control VR condition (Sham VR). REDCap Cloud (nPhase, Inc.) was used to apply an automatic and blinded randomization program and ensure equal allocation to both groups. Participants and study statisticians were blinded to treatment group assignment. Statisticians performed blinded analysis of data sets that were randomly labeled group A and group B with statistician unblinding occurring only after posttreatment month 3 data were collected and the data set locked (posttreatment month 3 data not yet analyzed). The 2 study coordinators (LG and IM) unblinded to individual treatment group assignments were not involved in any data analyses. Study participants remain blinded to treatment group assignment until their participation in the larger study is completed (8.5 months after randomization). The larger study quantifies long-term outcomes for the current treatment study; the study protocol and details are published elsewhere [43].

Study Interventions, VR Headset, and Software

Participants in both treatment groups (EaseVRx and Sham VR) received a Pico G2 4K all-in-one head-mounted VR device at no cost through postal mail. The Pico G2 4K device was used because they are commercially available, widely used, inexpensive, have minimal visual latency, and are easier for participants to use than many other devices. This hardware allows for displaying 3D images (EaseVRx) and 2D images (Sham VR). While each VR device contained software specific to the individual participant's assigned VR treatment group, all device packaging and directions for use were common to both treatment groups. Participants in this study were provided with online access to instructional materials outlining general use and set up of the headset. Relevant to the EaseVRx group, user exhalation is measured by the microphone embedded in the Pico G2 hardware, offering biodata-enabled immersive therapeutics. Participants in both groups were instructed to complete 1 VR program session daily for 56 days. Study staff monitored participant completion of the twice-weekly surveys and device use. Study staff provided reminders to complete surveys and otherwise were available upon request for technical support. The sections below describe the elements of the study interventions.

Therapeutic VR (EaseVRx)

Participants randomized and allocated to this treatment group received an immersive multimodal, skills-based, pain self-management VR program, called EaseVRx (AppliedVR), that incorporates evidence-based principles of CBT, mindfulness, and pain neuroscience education. The program content trains users on evidence-based pain and stress management strategies via immersive and enhanced biofeedback experiences. EaseVRx combines biopsychosocial education, diaphragmatic breathing training, relaxation response exercises that activate the parasympathetic nervous system, and executive functioning games to provide a mind–body approach toward living better with chronic pain. The standardized 56-day program

delivers a multifaceted combination of pain relief skills training through a prescribed sequence of daily immersive experiences. Each VR experience is 2-16 minutes in length (average of 6 minutes). The VR treatment modules were designed to minimize triggers of emotional distress or cybersickness. Treatment module categories included:

- Pain education: visual and voice-guided lessons establish a medical and scientific rationale for the VR exercises and behavioral medicine skills for pain relief.
- Relaxation/Interoception: scenes that progressively change from busy/active to calm in order to train users to understand the benefits of progressive relaxation.
- Mindful escapes: high-resolution 360 videos with therapeutic voiceovers, music, guided breathing, and sound effects designed to maximize the relaxation response and participant engagement.
- Pain distraction games: interactive games to train the skill of shifting focus away from pain.
- Dynamic breathing: breathing-based biofeedback training in immersive and interactive environments to support self-regulation and relaxation. These modules become increasingly challenging as users increase their skill with diaphragmatic breathing and parasympathetic control.

Sham VR

In compliance with VR-CORE clinical trial guidelines, we selected an active control that utilizes nonimmersive, 2D content within a VR headset as the most rigorous VR placebo [30]. The Sham VR headset displayed 2D nature footage (eg, wildlife in the savannah) with neutral music that was selected to be neither overly relaxing, aversive, nor distracting. The experience of Sham VR is similar to viewing nature scenes on a large-screen television and is not interactive. Twenty videos were rotated over the 56 sessions, with average duration of sessions closely matching those of EaseVRx (Figure 2).

Figure 2. Visual display of EaseVRx (skills-based, interactive, 3D) and Sham VR (non-interactive 2D nature scenes).

Research Standards and Compliance

In accordance with the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommendations, we included multiple measures to evaluate the importance of change in outcomes across 4 recommended domains: pain intensity, health-related quality of life and functioning, and ratings of overall improvement [46-48]. Additionally, measures and individual items were included to align directly with the National Institutes of Health (NIH) Pain Consortium's *Report on Research Standards for Chronic Low Back Pain* [49] or assess the domains recommended in the report. The study was performed in accordance with the CONSORT (Consolidated Standards of Reporting Trials) guidelines [50] (see Multimedia Appendix 2 for the completed checklist) and the recommended extension for reporting of psychological trials [51]. The Western Institutional Review Board approved the study (Puyallup, WA).

Data Collection and Time Points

Data collection included electronic participant-reported measures and objective VR device use data collected from the VR devices.

Data were collected across 3 phases of the study: pretreatment (days –14 to 0), active treatment (days 1-55), and end of treatment (day 56). The 14-day pretreatment phase involved administering the pain surveys 5 times (baseline, days –10, –7, –3, and 0); these measures were averaged within participants to establish a single pretreatment score for each variable assessed. During the 8-week active treatment phase, surveys were distributed biweekly (15 total surveys during treatment)

and at end of treatment on day 56. Accordingly, there were 17 time points per participant.

Study Variables and Measures

Table 1 outlines the timeline of variable assessment. This section details the measurement and methods used to assess each variable.

Table 1. Timeline of variable assessment.

Variables	Pretreatment phase (Days –14 to 0)		Treatment phase (Days 1 to 55)	End-of-treatment (Day 56)
	Baseline	Biweekly surveys	Biweekly surveys	Final assessment
Demographics and pain duration	X			
Pain intensity	X	X	X	X
Pain interference with activity, mood, sleep, stress	X	X	X	X
Patient's Global Impression of Change				X
Physical function	X			X
Sleep disturbance	X			X
Pain self-efficacy	X			X
Pain catastrophizing	X			X
Chronic pain acceptance	X			X
Prescription opioid use	X			X[a]
Over-the-counter analgesic medication use				X
Motion sickness and nausea				X[a]
Treatment satisfaction				X[a]
Virtual reality device use			X	X
System usability				X

[a]Because of a system error, these data were not captured at Day 56 as intended but at 1-month posttreatment.

Demographics and Pain Duration

Demographic variables included age, gender, level of education, race, ethnicity, employment status, annual household income, relationship status, duration of back pain (years since onset), state of residence, and zip code. To perform geospatial coding, rural–urban commuting area (RUCA) codes were downloaded from a public data set provided by the United States Department of Agriculture Economic Research Service [52]. Participants were classified as rural or urban based on their zip code. Finally, duration of time since pain onset was assessed.

Average Pain Intensity

The Defense and Veterans Pain Rating Scale (DVPRS) [53] was used to measure average pain intensity over the previous 24 hours using an 11-point numeric rating scale (0=no pain; 10=as bad as it could be and nothing else matters). Average pain intensity was assessed at baseline, pretreatment, during treatment, and at end of treatment on day 56.

Pain Interference With Activity, Mood, Sleep, and Stress

The DVPRS interference scale (DVPRS-II) was used to measure pain interference with activity, sleep, mood, and stress over the past 24 hours [54] (0=does not interfere; 10=completely interferes). Pain interference was assessed at baseline, pretreatment, during treatment, and at end of treatment on day 56.

Patient's Global Impression of Change

Aligning with IMMPACT recommendations for pain research [47], Patient's Global Impression of Change (PGIC) was assessed on day 56 (end of treatment) using the question, "Since the beginning of VR treatment, how would you describe the changes (if any) in activity limitations, symptoms, emotions and overall quality of life related to your low back pain?" on a 7-point scale ranging from 1 (No change or condition is worse) to 7 (A great deal better, and a considerable improvement that has made all the difference).

Physical Function and Sleep Disturbance (PROMIS)

The NIH Physical Function and Sleep Disturbance (PROMIS) [55] short-form measures were used to assess physical function (version 6b) [56] and sleep disturbance (version 6a) [57] over the past 7 days. Higher scores on physical function signify greater function whereas higher scores for sleep disturbance reflect greater symptom severity. The conversion table within the scoring manuals, made available from the Person-Centered Assessment Resource [58,59], was used to calculate the individual short-form T scores using the Item Response Theory scoring algorithms. Specifically, based on published item parameters, T scores (latent trait estimates) are computed for

each individual's response pattern using the Bayesian expected a posteriori method [60-62]. Widely applied in pain research [63-65], these measures were administered at baseline and posttreatment day 56.

Pain Catastrophizing

The 13-item Pain Catastrophizing Scale (PCS) [66] is a validated instrument widely used clinically and in pain research to assess patterns of negative cognition and emotion in the context of actual or anticipated pain. Despite having discrete subscales for rumination, magnification, and feelings of helplessness related to pain, prior work has shown that the PCS operates unidimensionally [67] (Cook et al, unpublished data). Aligning with prior work [15] and the goal of brevity, the following 4 PCS items were used: "It's terrible and I think it's never going to get any better," "I become afraid that the pain will get worse," "I can't seem to keep it out of my mind," and "I keep thinking about how badly I want the pain to stop." Respondents rate the frequency with which they experience such thoughts on a scale from 0 (Not at all) to 4 (All the time). Scores for the 4 items were summed to create a total score and index for pain catastrophizing. This measure was administered at baseline and on day 56.

Pain Self-Efficacy

Pain self-efficacy was assessed generally and also within the context of VR. For general pain self-efficacy, the 2-item Pain Self-Efficacy Questionnaire (PSEQ-2) is a validated instrument used to assess respondents' confidence in their ability to engage in various daily activities despite their chronic pain [68]. The PSEQ-2 comprises the following 2 items: "I can still accomplish most of my goals in life, despite the pain," and "I can live a normal lifestyle, despite the pain." Respondents use a 5-point scale to rate their response from 0 (Not at all confident) to 4 (Completely confident). Scores for the 2 items are summed to create a total score. The PSEQ-2 was administered at baseline and on day 56. For pain self-efficacy with a VR referent, at baseline, participants rated their overall confidence in their ability to manage their pain on a 10-point scale from 1 (Not at all Confident) to 10 (Very Confident). Following the intervention, this section will be divided into 2 items measuring their overall confidence levels while inside VR and outside VR.

Chronic Pain Acceptance

The 8-item Chronic Pain Acceptance Questionnaire (CPAQ-8) short form is an 8-item validated instrument that assesses one's engagement in personally meaningful activities despite pain, as well as efforts directed at controlling pain (example item: "I am getting on with the business of living no matter what my level of pain is") [69]. Respondents rate each item using a 6-point scale ranging from 0 (never true) to 5 (always true).

Satisfaction With Treatment

Satisfaction with treatment was assessed with several items. First, using a 6-point scale (0=strongly disagree and 5=strongly agree), participants rated 4 items: ease of use of the VR headset, enjoyment of the headset, whether the headset helped with pain coping, and desire to continue using the VR headset. These 4 items were summed to create a total satisfaction score. Additionally, 1 item assessed likelihood to recommend VR (0=definitely not recommend and 10=definitely would recommend). One item assessed likelihood to continue using VR if they were able to keep their headset using a response scale (0=definitely would not it and 10=definitely would use it). Because of an error with the electronic survey administration, these data were captured at 1 month posttreatment.

VR Device Use

Device use data were recorded by the devices (date and time stamped for device access and duration of use).

System Usability Scale

The System Usability Scale (SUS) is a validated, 10-item scale to assess a global view system usability (example item: "I thought the system was easy to use") [70]. Participants rate each item using a 5-point response scale ranging from 1=strongly disagree to 5=strongly agree. Some items are reverse scored, a multiplier is applied to the sum total, and total SUS scores range from 0-100.

Motion Sickness and Nausea (Cybersickness)

Adverse experiences with using VR was assessed using the question, "Did you experience any motion sickness or nausea while using VR?" on a 4-point scale, with 0=Never, 1=Sometimes, 2=Often, and 3=Always. Similar to prior work, cybersickness was assessed at the end of treatment [15]. Because of an error with the electronic survey administration, these data were captured at 1-month posttreatment.

Over-the-Counter Analgesic Medication Use

Participants were asked, "Do you take any 'over the counter' medication, meaning you can get yourself at a store without a prescription, to help you manage your back pain?" A binary response set (Y/N) was used to address variability in medication classes, formulations, doses, and frequency of use. Over-the-counter (OTC) analgesic use was measured at baseline and at posttreatment day 56.

Opioid Use Data

All opioid data were self-reported. Opioid medication doses were converted to a standardized morphine milligram equivalent daily dose using the Centers for Medicare & Medicaid Services "Opioid Oral Morphine Milligram Equivalent (MME) Conversion Table" [71]. Four assumptions were applied universally to all participants in calculating prescribed medication doses. First, participants who reported prescription medication use but did not report any of the classes of prescription medications were considered opioid free. Second, for participants who did not report the strength of their tablets, the most common dose of the tablet was used for the calculations (Hydrocodone 5 mg, Hydromorphone 2 mg, Oxycodone 5 mg, and Tramadol 50 mg); these doses are the lowest strength available for these medications and thus provide conservative estimates. Third, some participants reported using opioids "as needed" (ie, pro re nata [PRN] use) but did not detail their general frequency of use. For these cases, we calculated the dose and range based on 0 to maximum daily allowed (eg, for a participant prescribed medication every 6 hours PRN, we used the allowable range of 0-4 tablets per day and used the average value of 2 tablets per day). Fourth and last, if participants

reported their frequency of opioid use to be weekly, the reported dose was divided by 7 for calculating a daily MME; similarly, monthly reported doses were divided by 30 to calculate a daily MME.

Adverse Event Monitoring

Participants were provided with study staff contact information and encouraged to contact as needed and in the event of any problems using their device or with their treatment. Similar to other studies, cybersickness was assessed at the end of treatment [15]. However, due to a problem with the electronic survey, these data were not captured, and the survey was re-administered at 1-month posttreatment.

Sample Size Determination

A power analysis was performed using data from a prior RCT of a 21-day at-home VR for chronic pain compared to an audio-only version of the treatment [15]. This study revealed that an average pre–post treatment difference score in pain intensity was 1.48 for the VR group and 0.756 for the audio-only group (on an 11-point scale). Assuming an α level of .05 and 90% power, 45 participants per group would detect a treatment group × time interaction. To buffer against potential high attrition (40%), a minimum of 75 participants would be required per treatment group, with 90 participants per treatment group being ideal.

Statistical Analyses

All analyses involved 2-sided hypothesis tests, with α=.05 and adjusted for any multiple comparisons within the family of tests as appropriate. Group equivalence was assessed through univariate tests of association between treatment groups (EaseVRx/Sham VR) for all baseline demographic and clinical variables with chi-square and Kruskal–Wallis applied as appropriate.

The data were analyzed in a mixed-model framework (PROC GLIMMIX in SAS) using a marginal (population-averaged) model to allow for correlated responses across the repeated measures. There were 3 explanatory factors: treatment group, time, and time × treatment group. Treatment group, EaseVRx versus Sham VR, was specified as a fixed-effects factor. Time was specified as a random-effects factor to allow for correlated response using heterogeneous compound symmetry for the covariance structure within time. The 2 effects of interest were (1) the EaseVRx versus Sham VR between-group comparison across all time points, and (2) the time × treatment group effect which tests whether the treatment group influenced the trajectory of the key variables over time.

Data were 95.05% complete; missing values were not imputed for estimation of effects, but the predicted means were used in the graphical description. The primary outcomes were the time course of DVPRS (pain scale) from baseline (defined as the average of 3 pain ratings obtained during the 2 weeks before enrollment/randomization), at 8 weekly time points (twice per week) across the 8-week treatment period, and immediately posttreatment (Day 56). A linear mixed model was used with the treatment group (EaseVRx versus Sham VR) as an independent groups factor (ie, a between-subjects factor) and time of measurement as a dependent groups factor (ie, a within-subjects factor). DVPRS-II measures were analyzed using the same approach. Effect sizes for the EaseVRx versus Sham VR for the between-groups comparison were calculated using the standardized mean difference version of Cohen *d* [72]. The effect sizes for the within-subjects comparison (baseline to immediately posttreatment completion at day 56) were computed by treatment group using d_{rm}, an adaptation of Cohen *d* to suit the repeated measures design [73].

Results

Study Participants

Figure 3 shows the CONSORT diagram for the study (see Multimedia Appendix 2 for the CONSORT checklist). In total, 1577 individuals were assessed for eligibility and 1389 were excluded with the primary reason of having met the threshold for depressive symptoms. A total of 188 individuals were enrolled, randomized, and allocated to the treatment group. After randomization, 9 individuals discontinued participation, 5 were unable to receive a VR device, 1 returned their unopened device due to a recent medical diagnosis, and 4 voluntarily withdrew for unknown reasons. A total of 179 individuals received a VR device with their assigned treatment (EaseVRx [n=89] or Sham VR [n=90]). Because intention-to-treat analyses were performed, the analytic data set includes 11 individuals who did not provide complete data. Nearly 94% of participants in group A (84/89) and 93% of participants in group B (84/90) completed the day 56 assessment (end of treatment).

Figure 3. CONSORT Flow Diagram.

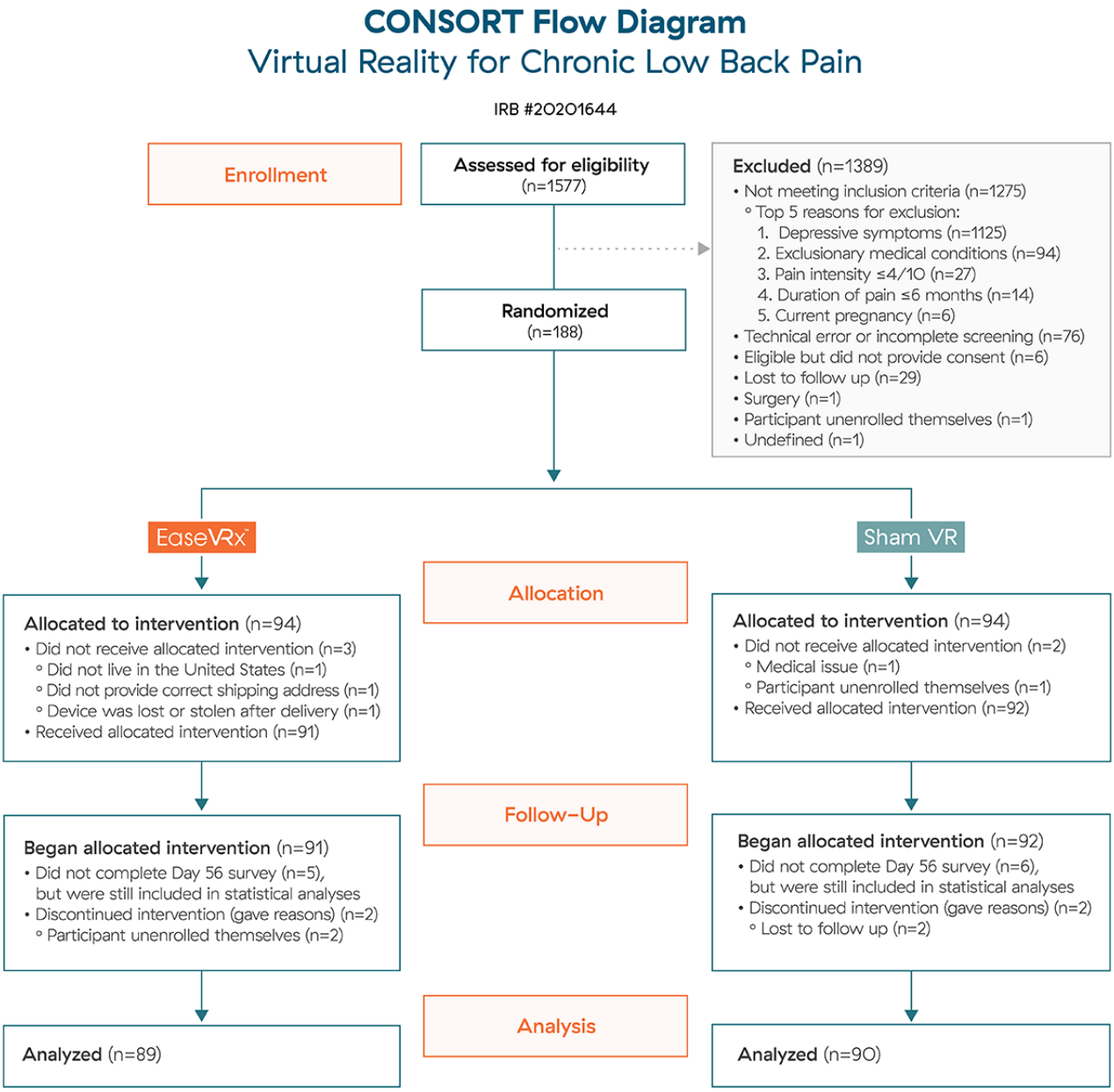


Table 2 displays the baseline demographic characteristics by treatment group. The sample included 179 participants from 40 states. RUCA codes were applied to categorize participants by zip code. In total, 76.5% (n=137) resided in highly urban or metropolitan areas, 13.4% (n=24) participants resided in metropolitan or micropolitan areas, 9.5% (n=17) were from small town or rural areas, and 0.6% (n=1) had no rural–urban identifier information. The sample was predominantly female (76.5%, 137/179), Caucasian (90.5%, 162/179), with at least some college education (91.1%, 163/179), and a mean age of 51.5 years (SD 13.1; range 18-81). No significant between-group differences were observed for any demographic variable, thus demonstrating that randomization was effective.

Table 3 presents the baseline pain and clinical characteristics for the sample by treatment group. The sample duration of back pain was 5 or more years, and the mean pain intensity score was 5/10 (SD 1.2; range 1-9). No significant differences were observed between treatment groups for all variables assessed.

Table 2. Baseline demographic characteristics by treatment group.

Demographics	Treatment group		
	EaseVRx (N=89)	Sham VR[a] (N=90)	*P*-value
Gender, n (%)			.527[b]
Male	22 (25)	19 (21)	
Female	67 (75)	70 (78)	
Other	0 (0)	1 (1)	
Age (years)			.964[c]
Mean (SD)	51.5 (13.5)	51.4 (12.9)	
Range	18.0-81.0	25.0-81.0	
Median (IQR)	51.0 (40.0-62.0)	54.0 (41.0-62.0)	
Race, n (%)			.247[b]
Asian	2 (2)	1 (1)	
Caucasian	78 (88)	84 (93)	
African American	5 (6)	1 (1)	
Multiracial	2 (2)	3 (3)	
Other	2 (2)	0 (0)	
Missing	0 (0)	1 (1)	
Education, n (%)			.142[b]
High-school graduate	6 (7)	9 (10)	
Some college	21 (24)	17 (19)	
Associate	10 (11)	16 (18)	
Undergraduate	17 (19)	25 (28)	
Postgraduate	35 (39)	22 (24)	
Missing	0 (0)	1 (1)	
Employment status, n (%)			.781[b]
Part-time	9 (10)	7 (8)	
Full-time	37 (42)	34 (38)	
Not working	13 (15)	10 (11)	
Retired	15 (17)	20 (22)	
Unable to work	15 (17)	18 (20)	
Missing	0 (0)	1 (1)	
Annual household income, n (%)			.665[b]
<US $40,000	22 (25)	22 (24)	
US $40,000-US $59,999	24 (27)	18 (20)	
US $60,000-US $79,999	16 (18)	18 (20)	
≥US $80,000	26 (29)	32 (36)	
Missing	1 (1)	0 (0)	
Marital status, n (%)			.605[b]
Married/Civil union	52 (58)	61 (68)	
Divorced/Widowed/Separated	20 (22)	14 (16)	
Single	10 (11)	10 (11)	

Demographics	Treatment group		
	EaseVRx (N=89)	Sham VR[a] (N=90)	*P*-value
Cohabitating	6 (7)	5 (6)	
Missing	1 (1)	0 (0)	

[a]VR: virtual reality.

[b]Chi-square *P*-value.

[c]Kruskal–Wallis *P*-value.

Table 3. Baseline clinical variables by treatment group.[a]

Variables	Treatment group		
	EaseVRx (N=89)	Sham VR[b] (N=90)	*P*-value
Pain duration, n (%)			.082[c]
<1 year	7 (8)	1 (1)	
1 year to <5 years	25 (28)	26 (29)	
5 years to <10 years	15 (17)	24 (27)	
>10 years	42 (47)	39 (43)	
Average pain intensity			.616[d]
Mean (SD)	5.1 (1.2)	5.2 (1.1)	
Range	2.2-8.2	2.8-7.8	
Median (IQR)	5.0 (4.2-5.8)	5.2 (4.4-5.6)	
Pain interference with activity			.398[d]
Mean (SD)	5.3 (1.8)	5.5 (1.5)	
Range	1.2-10.0	1.0-8.8	
Median (IQR)	5.6 (4.0-6.4)	5.5 (4.6-6.2)	
Pain interference with mood			.340[d]
Mean (SD)	4.5 (2.1)	4.7 (2.0)	
Range	0.0-8.8	0.2-9.6	
Median (IQR)	4.4 (2.8-5.8)	4.6 (3.6-5.8)	
Pain interference with sleep			.281[d]
Mean (SD)	4.8 (2.6)	5.3 (1.9)	
Range	0.0-10.0	0.6-9.6	
Median (IQR)	5.0 (3.0-7.0)	5.4 (3.8-6.4)	
Pain interference with stress			.852[d]
Mean (SD)	4.6 (2.2)	4.8 (2.0)	
Range	0.0-10.0	0.6-9.6	
Median (IQR)	4.8 (3.0-6.4)	5.0 (3.4-6.2)	
Pain self-efficacy			.766[d]
Mean (SD)	3.0 (1.5)	3.0 (1.2)	
Range	0.0-6.0	0.0-6.0	
Median (IQR)	3.0 (2.0-4.0)	3.0 (2.5-4.0)	
Pain catastrophizing			.977[d]
Mean (SD)	8.0 (3.8)	8.0 (3.5)	
Range	0.0-16.0	0.0-16.0	
Median (IQR)	8.0 (5.0-11.0)	7.0 (5.0-11.0)	
Physical function			.276[d]
Mean (SD)	38.3 (5.1)	37.6 (4.6)	
Range	21.0-48.9	27.1-59.0	
Median (IQR)	37.6 (35.0-41.2)	37.6 (35.0-40.2)	
Chronic pain acceptance			.708[d]
Mean (SD)	24.5 (7.3)	23.9 (6.7)	

Variables	Treatment group		
	EaseVRx (N=89)	Sham VR[b] (N=90)	*P*-value
Range	5.0-42.0	7.0-47.0	
Median (IQR)	24.0 (20.0-28.0)	23.0 (20.0-28.0)	
Sleep disturbance			.230[d]
Mean (SD)	56.7 (5.2)	57.6 (4.4)	
Range	44.2-67.5	45.5-69.0	
Median (IQR)	56.3 (53.3-60.4)	58.3 (55.3-60.4)	
Opioid use, n (%)	22 (25)	33 (37)	.083
Opioid dose (daily morphine milligram equivalent)			.158[d]
Mean (SD)	25.2 (106.2)	15.3 (41.1)	
Range	0.0-875.0	0.0-300.0	
Median (IQR)	0.0 (0.0-0.0)	0.0 (0.0-10.0)	
OTC analgesic use, n (%)	61 (69)	55 (61)	.320

[a]Baseline for the 5 pain variables (pain intensity, pain-related activity, mood, sleep, and stress interference) represents the average from 5 administrations in the pretreatment phase (days –14, –10, –7, –3, and 0).

[b]VR: virtual reality.

[c]Chi-square *P*-value.

[d]Kruskal–Wallis *P*-value.

Treatment Engagement

Device use data revealed nonsignificant between-group differences for treatment engagement: EaseVRx participants completed a total of 43.30 (SD 15.91) experiences (average 5.4 per week) and Sham VR participants completed 48.06 (SD 24.78) experiences (average 6.0 per week).

Device Safety and Adverse Events

Of the 147 participants who completed the 1-month posttreatment survey, 7/72 (9.7%) from the EaseVRx group and 5/75 (6.7%) from the Sham VR group reported experiencing nausea and motion sickness during the treatment phase of the study (*P*=.50). Participants were encouraged to contact study staff with any problems experienced during treatment; however, no participants contacted study staff to report adverse events of any type, including nausea and motion sickness.

Primary Outcomes

Common analyses and data visualization were applied for all primary outcomes. The x-axis represents time (days), with days –14 to 0 averaged and labeled “day 7” to represent the pretreatment phase, days 1-55 were the active treatment phase, and day 56 was the end of treatment and the primary endpoint. The color bands represent the 95% CI values for the mean after correcting for multiple comparisons (Tukey–Kramer). Overlapping bands indicate nonsignificant treatment group differences (*P*-value) of simple main effects within each time point. The corresponding model effects for each primary outcome are displayed in Figures 4-8.

Figure 4. Average pain intensity.

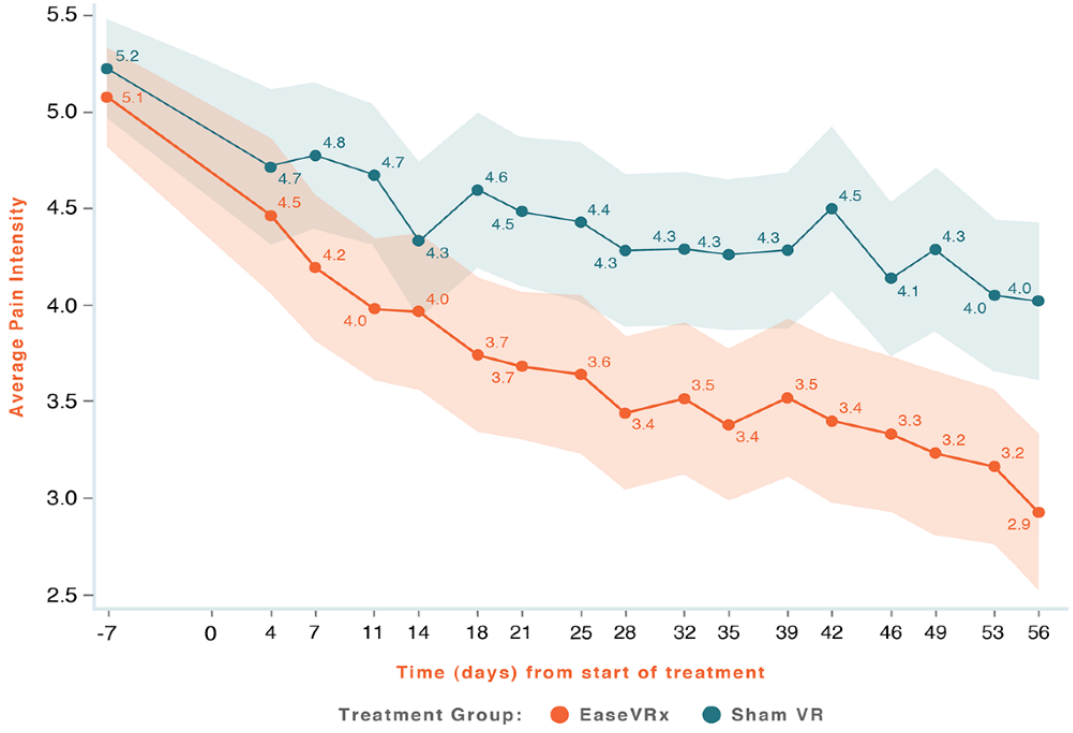


Figure 5. Pain-Related Interference with Activity.

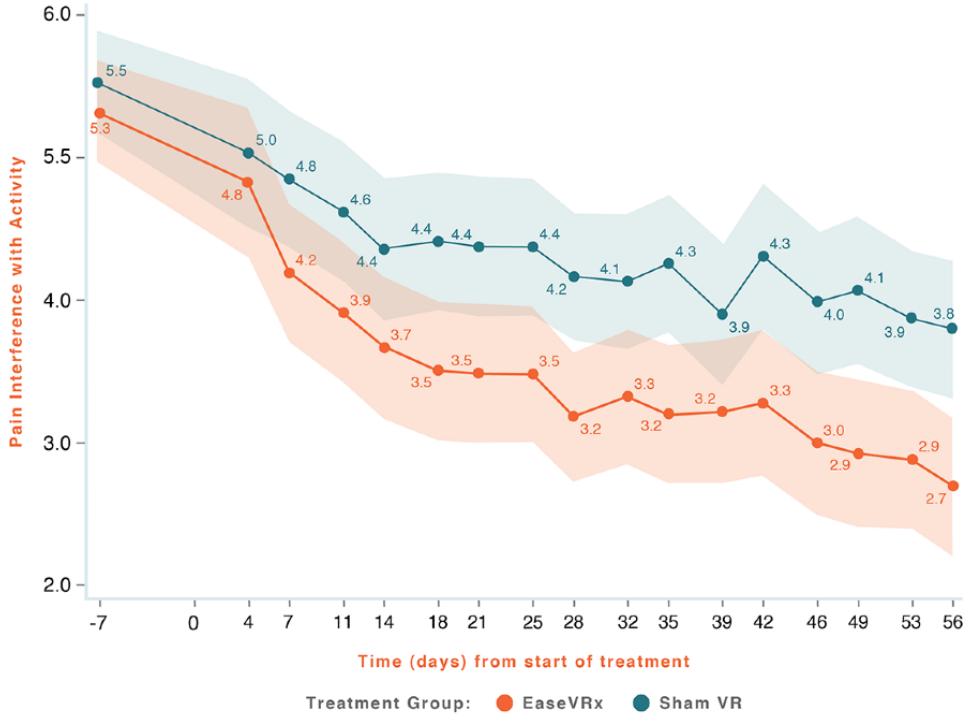


Figure 6. Pain-Related Interference with Mood.

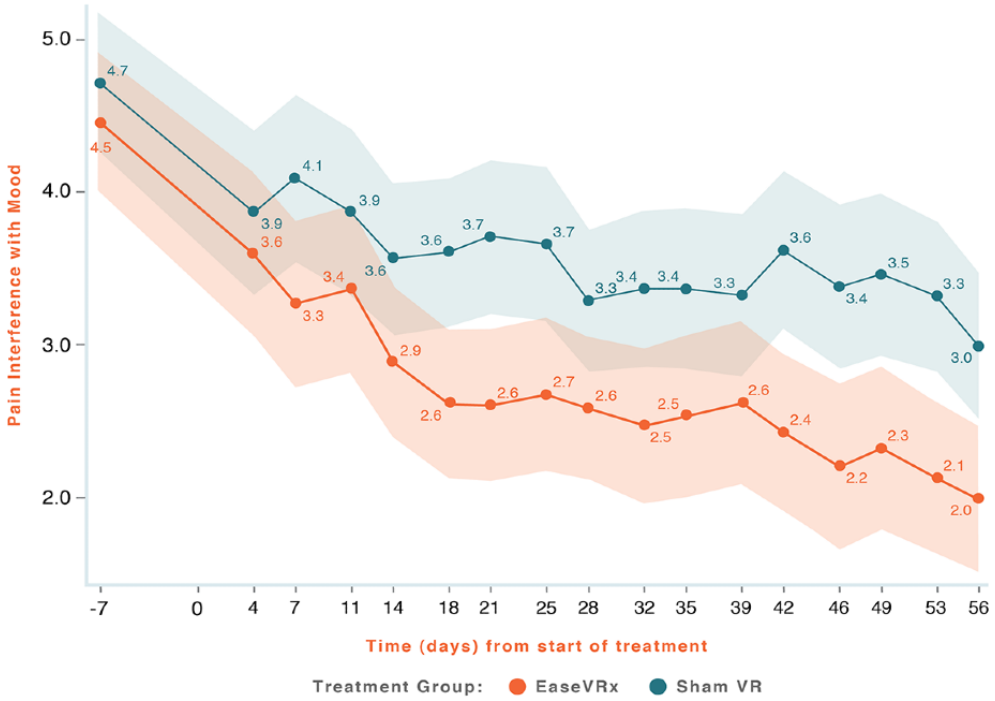


Figure 7. Pain Related Interference with Sleep.

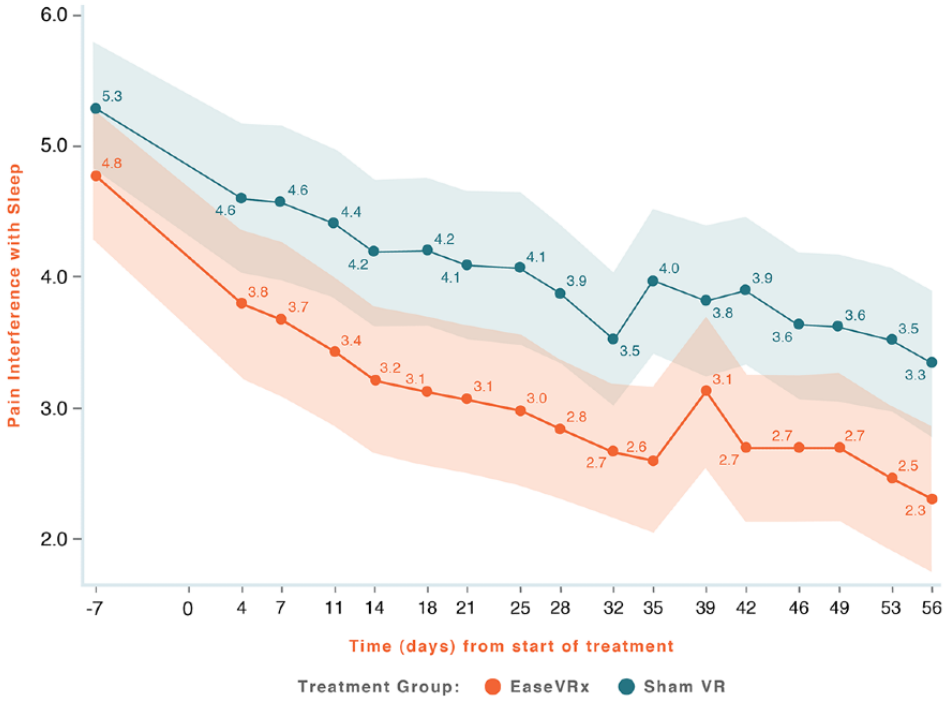
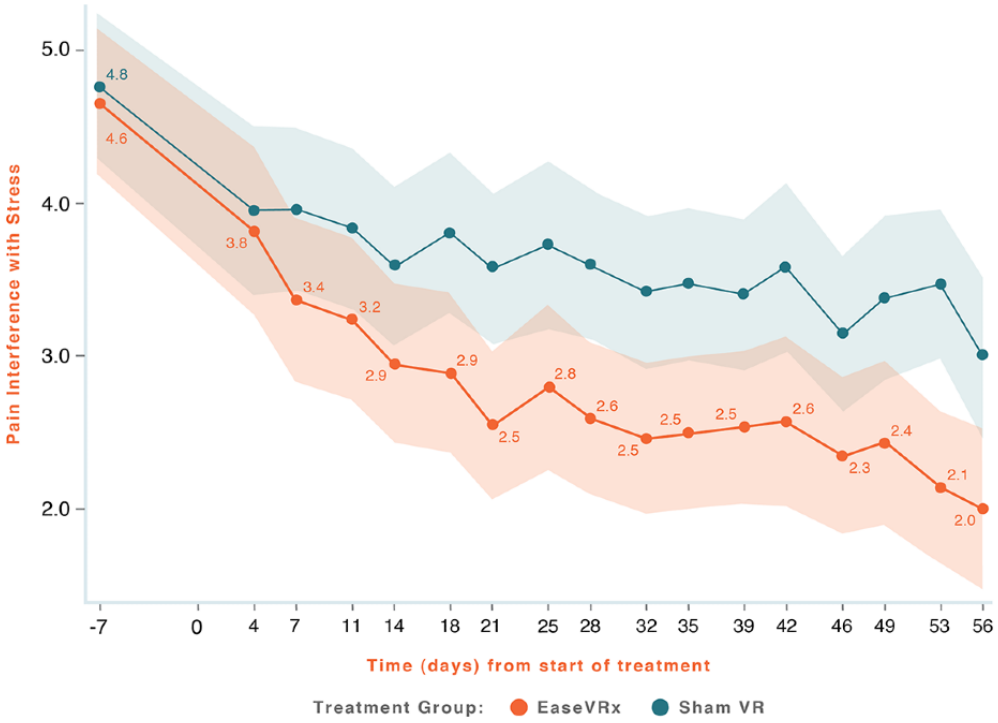


Figure 8. Pain-Related Interference with Stress.



We observed a significant treatment effect (*P*=.001); on average, the EaseVRx group had lower pain intensity compared to the Sham VR group (Cohen *d*=0.49). Separately, we observed a time effect; average pain intensity significantly decreased over time for both treatment groups (time effect, *P*<.001). Most importantly, the decrease was greater for EaseVRx versus Sham VR (treatment × time effect, *P*<.001). Pain intensity reduced by an average of 42.8% for the EaseVRx group and 25.1% for the Sham VR group. The d_{rm} for EaseVRx was 1.31, with combined results showing large effect size and moderate clinical importance. The VR Sham group d_{rm} was 0.75, with combined results showing a large effect size and minimal clinical importance. As much as 65% (55/84) of EaseVRx and 40% (34/84) of Sham VR participants achieved 30% or more reduction in pain intensity. For EaseVRx, 46% (39/84) achieved 50% or more pain reduction, while for Sham VR 26% (22/84) reached that threshold.

We observed a significant treatment effect (*P*=.004) on pain interference with activity; on average, the EaseVRx group had lower activity interference compared to the Sham VR group (Cohen *d*=0.44). We also observed a time effect; pain interference with activity decreased over time for both treatment groups (time effect, *P*<.001). Most importantly, the decrease was greater for EaseVRx versus Sham VR (treatment × time effect, *P*=.013). Pain interference with activity reduced by an average of 51.6% for the EaseVRx group and 32.4% for the Sham VR group. The d_{rm} for the EaseVRx group was 1.27, with combined results showing large effect size and moderate clinical importance. As much as 71% (60/84) of EaseVRx and 57% (48/84) of Sham VR participants achieved 30% or more reduction in pain-interference with activity, and 56% of (47/84) of the EaseVRx participants achieved 50% or more reduction. The VR Sham group d_{rm} was 0.97, with combined results showing a large effect size and moderate clinical importance.

We observed a significant treatment effect (*P*=.005) on pain interference with mood; on average, the EaseVRx group had lower mood interference compared to the Sham VR group (Cohen *d*=0.42). We also observed a time effect; pain-interference with mood decreased over time for both treatment groups (time effect, *P*<.001) and the decrease was greater for EaseVRx versus Sham VR (treatment × time effect, *P*=.010). Pain interference with mood reduced by an average of 55.7% for EaseVRx and 40.04% for the Sham VR. The d_{rm} for the EaseVRx was 1.18, with combined results evidencing a large effect size and substantial clinical importance. As much as 74% (62/84) of EaseVRx participants and 60% (50/84) of Sham VR participants achieved 30% or more reduction in pain-related interference with mood, and 61% (51/84) of the EaseVRx participants achieved 50% or more reduction. The d_{rm} for the VR Sham group was 0.79, with combined results showing a moderate effect size and moderate clinical importance.

We observed a significant treatment effect (*P*=.004) on pain-interference with sleep; on average, the EaseVRx group had lower sleep interference compared to the Sham VR group (Cohen *d*=0.44). We also observed a time effect; pain-interference with sleep decreased over time for both treatment groups (time effect, *P*<.001). However, there was no difference between treatment groups over time (*P*=.755). Pain interference with sleep reduced by an average of 54% for EaseVRx and 39.2% for the Sham VR. The d_{rm} for the EaseVRx was 0.95, with combined results showing large effect size and substantial clinical importance for symptom reduction. As much as 70% (59/84) of EaseVRx and 60% (50/84) of participants achieved 30% or more reduction in Pain-related interference with sleep, and 60% (50/84) of the EaseVRx participants achieved 50% or more reduction. The VR Sham group d_{rm} was

0.87, with combined results showing a large effect size and moderate clinical importance.

We observed a significant treatment effect (P=.009) on pain interference with stress; on average, the EaseVRx group had lower stress interference compared to the Sham VR group (Cohen d=0.40). We also observed a time effect; pain interference with stress decreased over time for both treatment groups (time effect, P<.001). Most importantly, the decrease was greater for EaseVRx versus Sham VR (treatment × time effect, P=.004). Pain interference with stress reduced by an average of 59.9% for the EaseVRx group and 38.3% for the Sham VR group. The d_{rm} for the EaseVRx was 1.17, with combined results evidencing a large effect size and substantial clinical importance. As much as 76% (64/84) of EaseVRx participants and 56% (47/84) of the Sham VR participants achieved 30% or more reduction in pain-related interference with stress, and 63% (53/84) achieved 50% or more reduction. The VR Sham group d_{rm} was 0.77, with combined results showing a large effect size and moderate clinical importance.

Secondary Outcomes

Treatment Engagement

Device use data were received for 149 participants (EaseVRx =77; Sham VR = 72). EaseVRx participants completed a mean of 43.3 (SD 15.9) sessions, while the Sham VR group completed a mean of 48.1 (SD 24.8) sessions. A significant group difference for treatment engagement was not found.

Patient's Global Impression of Change

Between-Subjects Analysis

The between-subjects analysis of PGIC at posttreatment indicated a significant effect of condition (P=.002); participants in the EaseVRx group reported greater PGIC than those in the Sham VR group (4.13 versus 3.11).

While both groups evidenced improvement in pain coping symptoms, including pain catastrophizing, pain self-efficacy, and pain acceptance from pretreatment to end of treatment, none of these improvements achieved statistical significance.

Physical Function

For PROMIS Physical Function, we observed a significant treatment effect (P=.022); the EaseVRx group had higher physical function compared to the Sham VR group (Cohen d=0.34). Both treatment groups significantly improved from baseline to postintervention; there was superior functional improvement for the EaseVRx group relative to the Sham VR group (time × condition effect, P=.002). The d_{rm} values for the EaseVRx and VR Sham groups were 0.64 and 0.35, respectively.

Sleep Disturbance

For PROMIS sleep disturbance, we observed a significant treatment effect (P=.013); the EaseVRx group had lower sleep disturbance compared to the Sham VR group (Cohen d=0.37). Both treatment groups significantly improved throughout the study; there was superior improvement for the EaseVRx group relative to the Sham VR group (time × condition effect, P=.035). The d_{rm} for the EaseVRx group was 0.83, evidencing a large effect and substantial clinical importance.

Prescription Opioid and OTC Analgesic Use

Neither treatment group evidenced a significant change in MME dose from baseline to end of treatment. For OTC analgesic medication use, a substantial decrease was observed in the EaseVRx group. While 61 reported using OTC analgesics at baseline, 50 reported use at posttreatment day 56 (P=.01). For Sham VR, 55 and 56 reported OTC analgesic use at baseline and posttreatment, respectively (nonsignificant).

Treatment Satisfaction, Likelihood to Recommend, and Likelihood to Continue Use

For the 4 summed satisfaction items, the EaseVRx group reported greater satisfaction with treatment than the Sham VR group (4.32 versus 3.46 respectively; P<.001). Similarly, the EaseVRx group reported greater likelihood to recommend VR to someone else compared to the Sham VR group (8.72 versus 6.55, respectively; P<.001). Finally, EaseVRx participants reported greater likelihood to continue using VR if they could keep their headset compared to Sham VR (9.18 versus 7.23, respectively; P<.001).

VR Usability Ratings

Both treatment groups reported high usability with no statistical difference between groups (EaseVRx usability rating = 84.33; Sham VR usability rating = 81.16).

Additional Analyses

Two additional Sham VR participants provided data for only 1 and 2 of the 16 surveys; including them in the analysis (for a total of 92) did not alter the significance of the study findings for any variable at any time point.

Discussion

Principal Findings

We conducted the first placebo-controlled RCT of home-based therapeutic VR in a national sample of individuals with cLBP. We hypothesized that an 8-week pain relief skills VR program (EaseVRx) would be superior to Sham VR at posttreatment (day 56) for our primary outcomes: average pain intensity and pain-related interference with activity, mood, sleep, and stress. While both study groups had significant reductions in pain and all domains of pain-related interference, EaseVRx evidenced superior treatment effects for all primary outcomes except sleep interference; the between-groups Cohen d effect sizes ranged from 0.40 to 0.49. For EaseVRx, large pre/posttreatment Cohen d effect sizes ranged from 1.17 to 1.3 and demonstrated moderate to substantial clinical importance for reduced pain intensity and pain-related interference with activity, mood, and stress at end of treatment. A greater proportion of participants in the EaseVRx group exceeded thresholds for clinical importance of effects. For moderate clinical importance in pain reduction (≥30% reduction in pain), 65% (55 of 84) in EaseVRx versus 40% (34 of 84) in Sham VR met this threshold. For substantial clinical importance in pain reduction (≥50% reduction in pain), 46% (39 of 84) in EaseVRx versus 26% (22

of 84) in Sham VR met this threshold. These effects for therapeutic VR exceeded effects reported for a 3-week skills-based VR program for chronic pain [15] and effects reported for CBT studies involving in-person 8-week treatment with a trained therapist [4,6,7].

Both treatment groups evidenced moderate to substantial reductions for pain-related interference with sleep; however, no between group differences were found. By contrast, EaseVRx was superior to Sham VR for reducing sleep disturbance (secondary outcome), suggesting that therapeutic VR is particularly effective for reducing general versus pain-related sleep disturbance.

The secondary outcomes yielded interesting findings. First, physical function significantly improved for both treatment groups, with superior improvements found for EaseVRx. We note that the therapeutic content included no kinematic elements, nor did it include direction for activity, movement, or goal setting for either. Accordingly, improvement in physical function may be a product of substantially reduced pain-related interference in activity. Next, our hypothesis that therapeutic VR would be superior to Sham VR for improving pain coping (eg, reducing pain catastrophizing and improving pain self-efficacy and chronic pain acceptance) was unmet. The lack of effect for pain catastrophizing was particularly striking because this is a malleable construct that is highly responsive to behavioral treatments broadly [4,6,7] (Darnall et al, unpublished). However, our results align with the prior 3-week VR program study that similarly found no effect on pain catastrophizing and pain self-efficacy. Findings suggest that increased treatment time and additional focal content were insufficient to affect these factors significantly. These findings also highlight differences in treatment efficacy between therapeutic VR and CBT. Multiple studies have shown that CBT imparts its largest effect on pain catastrophizing [4]. By contrast, therapeutic VR evidenced its largest effects for reducing pain intensity and pain interference across several key domains.

We found no differences for either treatment group for change in prescription opioid use, and note that prescribing changes are unlikely to occur within the 2-month timeframe of the study. Similarly, we found no changes in “as needed” opioid use. We found substantially reduced use of OTC analgesic medication at posttreatment for the EaseVRx group only. While additional research is needed to replicate this finding, this finding offers a promising suggestion that therapeutic VR may reduce need for analgesics. Future VR research should examine medication use in greater detail and with higher-frequency data capture.

The study’s methodologic rigor was strengthened by a placebo treatment (Sham VR), which evidenced equivalent participant engagement as therapeutic VR. The extant literature on digital behavioral health research has reported participant treatment engagement rates ranging from 20% to 60% [13,74-76]. Strikingly, the current trial evidenced a 90% engagement rate in both groups, thus suggesting that efforts to enhance the face validity of the Sham VR were effective. These results also highlight the public interest in home-based VR as a chronic pain treatment modality. Therapeutic VR was rated significantly higher than Sham VR for satisfaction, likelihood to recommend to others, and likelihood to continue using the device after the 8-week treatment phase if it was made available. Combined with high participant engagement data and device usability ratings, these data extend prior work [15] supporting the utility, user satisfaction, and efficacy of home-based VR for chronic pain.

The context of the COVID-19 pandemic may have influenced the participant engagement rate. This trial occurred entirely at a time when people were adhering to strict social distancing measures and were environmentally isolated. Indeed, for many people receipt of medical care is worryingly low due to limited availability or unavailability of outpatient treatment options. These circumstances likely supported interest in effective home-based pain care. The COVID-19 context and the home-based study design support the ecological validity of the study findings. Notably, the study was conducted remotely and did not benefit from any in-person contacts or enhanced placebo effects that occur when research involves high-touch protocols or is conducted in medical treatment settings (ie, halo effects).

Strengths of this study include methods that attended to the IMMPACT recommendations and the NIH Research Standards for Back Pain. The study was conducted in a national sample drawn from 40 states, was well-distributed geographically, and included participants from urban and rural settings. Additional aspects of methodological rigor included participant and analyst blinding, intention-to-treat analyses, randomization, and a rigorous placebo control group that adhered to recommended specifications for an optimal VR sham [77].

Limitations

Several limitations bear consideration when evaluating the study results. With the exception of device use metrics, all data were self-reported. The study was untethered from medical care and thus, there was no ability to confirm pain diagnoses or analgesic prescription information. The study sample was predominantly female, white, college educated, and internet savvy; thus, findings may not generalize to individuals with disparate demographic characteristics. These findings are consistent with previous evidence showing that highly educated females are more likely to use self-care mobile health technologies, particularly those with mindfulness-based content [78], and this dovetails with a general female predilection to seek treatment for pain and other health concerns. As eHealth literacy and awareness increase in clinicians and the general population, it is likely that health technologies (including therapeutic VR) will benefit other demographics [79]. Additionally, the study was conducted in individuals with cLBP and findings may not generalize to other chronic pain conditions.

Data on cybersickness were collected at 1 month posttreatment and this lag introduces potential for recall bias, despite others documenting that participants readily recall cybersickness due to its specificity and salience [15]. Overall attrition was low (n=11) and it is possible that the 2 EaseVRx participants who left the study after receiving their headset did so due to cybersickness. No participants contacted study staff to report adverse events of any type. Data on sex differences for cybersickness are mixed, with some reporting a female

preponderance, while a recent meta-analysis suggests no sex effect [80]. Within the context of our limitations, we highlight low reports of cybersickness in a predominantly female sample. Future research may better capture potential VR adverse effects by assessing these factors in the first week of treatment. Interpretation of data on prescription opioid use was limited by low-frequency sampling methods that are subject to recall bias and poor data accuracy. Opioid prescriptions often allow "as needed" flexibilities in medication use and future research designs may benefit from high-frequency sampling methods which improve data accuracy. Further, in quantifying analgesic medication use (prescription opioids and OTC analgesics) we did not assess or control for life events or circumstances that may have influenced medication use (eg, acute injury or surgery). Finally, our threshold for depression screening and inclusion was applied to provide greater specificity [44] yet is noted to be lower than what is reported for many individuals with chronic pain. While our purpose in applying a low threshold for depressive symptoms was to minimize poor engagement and attrition resulting from anhedonia or avolition (and therefore poor data quality to determine treatment efficacy), recent research in cLBP suggests such concerns may be unfounded for mild to moderate depressive symptoms (Darnall et al, unpublished).

As a fully self-administered and on-demand treatment, VR is a promising and effective option that can transcend many traditional barriers to nonpharmacologic pain treatment; however, currently access is limited. With future reimbursement and commercial availability, therapeutic VR could become affordable and widely accessible to consumers.

Conclusion

An 8-week self-administered home-based pain relief skills VR program appears effective for reducing pain intensity and pain-related interference in activity, mood, and stress posttreatment. Treatment effects ranged from moderately to substantially clinically important. Therapeutic VR had high rates for engagement and user satisfaction. Additional studies are needed to determine effects in demographically diverse populations and in other pain conditions. Data suggest therapeutic VR is not operating through traditional pain coping mechanisms. As such, additional research is needed to characterize mechanisms of treatment effects and durability of effects. Home-based VR appears to provide effective and on-demand nonpharmacologic treatment for cLBP.

Acknowledgments

AppliedVR, Inc supported this study.

Conflicts of Interest

LG, TM, and IM are employees of AppliedVR, Inc. JS is president of AppliedVR, Inc. BD is chief science advisor for AppliedVR, Inc. BB, PK, and VS are consultants for AppliedVR, Inc. RL is a prior advisor and minor shareholder of AppliedVR, Inc.

Multimedia Appendix 1
eConsent form.
[PDF File (Adobe PDF File), 115 KB - jmir_v23i2e26292_app1.pdf]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V.1.6.1).
[PDF File (Adobe PDF File), 380 KB - jmir_v23i2e26292_app2.pdf]

References

1. Institute of Medicine (US) Committee on Advancing Pain Research, Care, and Education. Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research. Washington (DC): National Academies Press; 2011.
2. Freburger JK, Holmes GM, Agans RP, Jackman AM, Darter JD, Wallace AS, et al. The rising prevalence of chronic low back pain. Arch Intern Med 2009 Feb 09;169(3):251-258 [FREE Full text] [doi: 10.1001/archinternmed.2008.543] [Medline: 19204216]
3. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, Lancet Low Back Pain Series Working Group. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018 Jun 09;391(10137):2368-2383. [doi: 10.1016/S0140-6736(18)30489-6] [Medline: 29573872]
4. Williams A, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2012 Nov 14;11:CD007407 [FREE Full text] [doi: 10.1002/14651858.CD007407.pub3] [Medline: 23152245]
5. Medicaid Strategies for Non-Opioid Pharmacologic and NonPharmacologic Chronic Pain Management. 2019. URL: https://www.medicaid.gov/sites/default/files/federal-policy-guidance/downloads/cib022219.pdf [accessed 2021-02-03]
6. Turner JA, Anderson ML, Balderson BH, Cook AJ, Sherman KJ, Cherkin DC. Mindfulness-based stress reduction and cognitive behavioral therapy for chronic low back pain: similar effects on mindfulness, catastrophizing, self-efficacy, and

acceptance in a randomized controlled trial. Pain 2016 Nov 30;157(11):2434-2444 [FREE Full text] [doi: 10.1097/j.pain.0000000000000635] [Medline: 27257859]
7. Cherkin DC, Sherman KJ, Balderson BH, Cook AJ, Anderson ML, Hawkes RJ, et al. Effect of Mindfulness-Based Stress Reduction vs Cognitive Behavioral Therapy or Usual Care on Back Pain and Functional Limitations in Adults With Chronic Low Back Pain: A Randomized Clinical Trial. JAMA 2016 Mar 22;315(12):1240-1249 [FREE Full text] [doi: 10.1001/jama.2016.2323] [Medline: 27002445]
8. Darnall BD, Scheman J, Davin S, Burns JW, Murphy JL, Wilson AC, et al. Pain Psychology: A Global Needs Assessment and National Call to Action. Pain Med 2016 Feb 23;17(2):250-263 [FREE Full text] [doi: 10.1093/pm/pnv095] [Medline: 26803844]
9. Mojtabai R. National trends in long-term use of prescription opioids. Pharmacoepidemiol Drug Saf 2018 May 06;27(5):526-534. [doi: 10.1002/pds.4278] [Medline: 28879660]
10. Eccleston C, Blyth FM, Dear BF, Fisher EA, Keefe FJ, Lynch ME, et al. Managing patients with chronic pain during the COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services. Pain 2020 May 2;161(5):889-893 [FREE Full text] [doi: 10.1097/j.pain.0000000000001885] [Medline: 32251203]
11. Fisher E, Law E, Palermo TM, Eccleston C. Psychological therapies (remotely delivered) for the management of chronic and recurrent pain in children and adolescents. Cochrane Database Syst Rev 2014 May 14;2014:CD011118 [FREE Full text] [doi: 10.1002/14651858.CD011118] [Medline: 25221436]
12. Ziadni M, Gonzalez-Castro LA, Darnall BD. Virtual Single-Session. 2020. URL: https://clinicaltrials.gov/ct2/show/NCT04546685 [accessed 2021-02-03]
13. Darnall B, Ziadni MS, Krishnamurthy P, Flood P, Heathcote LC, Mackey IG, et al. "My Surgical Success": Effect of a Digital Behavioral Pain Medicine Intervention on Time to Opioid Cessation After Breast Cancer Surgery-A Pilot Randomized Controlled Clinical Trial. Pain Med 2019 Nov 01;20(11):2228-2237 [FREE Full text] [doi: 10.1093/pm/pnz094] [Medline: 31087093]
14. Pain Trainer. URL: https://www.paintrainer.org/ [accessed 2021-02-03]
15. Darnall BD, Krishnamurthy P, Tsuei J, Minor JD. Self-Administered Skills-Based Virtual Reality Intervention for Chronic Pain: Randomized Controlled Pilot Study. JMIR Form Res 2020 Jul 07;4(7):e17293 [FREE Full text] [doi: 10.2196/17293] [Medline: 32374272]
16. Navarro-Haro MV, Modrego-Alarcón M, Hoffman HG, López-Montoyo A, Navarro-Gil M, Montero-Marin J, et al. Evaluation of a Mindfulness-Based Intervention With and Without Virtual Reality Dialectical Behavior Therapy Mindfulness Skills Training for the Treatment of Generalized Anxiety Disorder in Primary Care: A Pilot Study. Front Psychol 2019 Jan 28;10:55 [FREE Full text] [doi: 10.3389/fpsyg.2019.00055] [Medline: 30745888]
17. Lindner P, Dagöö J, Hamilton W, Miloff A, Andersson G, Schill A, et al. Virtual Reality exposure therapy for public speaking anxiety in routine care: a single-subject effectiveness trial. Cogn Behav Ther 2021 Jan 01;50(1):67-87. [doi: 10.1080/16506073.2020.1795240] [Medline: 32870126]
18. Miloff A, Lindner P, Dafgård P, Deak S, Garke M, Hamilton W, et al. Automated virtual reality exposure therapy for spider phobia vs. in-vivo one-session treatment: A randomized non-inferiority trial. Behav Res Ther 2019 Jul;118:130-140. [doi: 10.1016/j.brat.2019.04.004] [Medline: 31075675]
19. Tejera D, Beltran-Alacreu H, Cano-de-la-Cuerda R, Leon Hernández JV, Martín-Pintado-Zugasti A, Calvo-Lobo C, et al. Effects of Virtual Reality versus Exercise on Pain, Functional, Somatosensory and Psychosocial Outcomes in Patients with Non-specific Chronic Neck Pain: A Randomized Clinical Trial. Int J Environ Res Public Health 2020 Aug 16;17(16):5950 [FREE Full text] [doi: 10.3390/ijerph17165950] [Medline: 32824394]
20. Hoffman HG, Richards T, Coda B, Richards A, Sharar SR. The illusion of presence in immersive virtual reality during an fMRI brain scan. Cyberpsychol Behav 2003 Apr;6(2):127-131. [doi: 10.1089/109493103321640310] [Medline: 12804024]
21. Gold JI, Belmont KA, Thomas DA. The neurobiology of virtual reality pain attenuation. Cyberpsychol Behav 2007 Aug;10(4):536-544. [doi: 10.1089/cpb.2007.9993] [Medline: 17711362]
22. Smith V, Warty RR, Sursas JA, Payne O, Nair A, Krishnan S, et al. The Effectiveness of Virtual Reality in Managing Acute Pain and Anxiety for Medical Inpatients: Systematic Review. J Med Internet Res 2020 Nov 02;22(11):e17980 [FREE Full text] [doi: 10.2196/17980] [Medline: 33136055]
23. Gold J, Mahrer NE. Is Virtual Reality Ready for Prime Time in the Medical Space? A Randomized Control Trial of Pediatric Virtual Reality for Acute Procedural Pain Management. J Pediatr Psychol 2018 Apr 01;43(3):266-275. [doi: 10.1093/jpepsy/jsx129] [Medline: 29053848]
24. Atzori B, Hoffman HG, Vagnoli L, Patterson DR, Alhalabi W, Messeri A, et al. Virtual Reality Analgesia During Venipuncture in Pediatric Patients With Onco-Hematological Diseases. Front Psychol 2018 Dec 20;9:2508 [FREE Full text] [doi: 10.3389/fpsyg.2018.02508] [Medline: 30618938]
25. Atzori B, Lauro Grotto R, Giugni A, Calabrò M, Alhalabi W, Hoffman HG. Virtual Reality Analgesia for Pediatric Dental Patients. Front Psychol 2018 Nov 23;9:2265 [FREE Full text] [doi: 10.3389/fpsyg.2018.02265] [Medline: 30532720]
26. Gold JI, Kim SH, Kant AJ, Joseph MH, Rizzo A. Effectiveness of virtual reality for pediatric pain distraction during i.v. placement. Cyberpsychol Behav 2006 Apr;9(2):207-212. [doi: 10.1089/cpb.2006.9.207] [Medline: 16640481]

27. Haisley KR, Straw OJ, Müller DT, Antiporda MA, Zihni AM, Reavis KM, et al. Feasibility of implementing a virtual reality program as an adjuvant tool for peri-operative pain control; Results of a randomized controlled trial in minimally invasive foregut surgery. Complement Ther Med 2020 Mar;49:102356. [doi: 10.1016/j.ctim.2020.102356] [Medline: 32147069]
28. Koç Özkan T, Polat F. The Effect of Virtual Reality and Kaleidoscope on Pain and Anxiety Levels During Venipuncture in Children. J Perianesth Nurs 2020 Apr;35(2):206-211. [doi: 10.1016/j.jopan.2019.08.010] [Medline: 31759833]
29. Hoffman HG, Chambers GT, Meyer WJ, Arceneaux LL, Russell WJ, Seibel EJ, et al. Virtual reality as an adjunctive non-pharmacologic analgesic for acute burn pain during medical procedures. Ann Behav Med 2011 Apr 25;41(2):183-191 [FREE Full text] [doi: 10.1007/s12160-010-9248-7] [Medline: 21264690]
30. Khadra C, Ballard A, Déry J, Paquin D, Fortin J, Perreault I, et al. Projector-based virtual reality dome environment for procedural pain and anxiety in young children with burn injuries: a pilot study. JPR 2018 Feb;Volume 11:343-353. [doi: 10.2147/jpr.s151084]
31. Spiegel B, Fuller G, Lopez M, Dupuy T, Noah B, Howard A, et al. Virtual reality for management of pain in hospitalized patients: A randomized comparative effectiveness trial. PLoS One 2019 Aug 14;14(8):e0219115 [FREE Full text] [doi: 10.1371/journal.pone.0219115] [Medline: 31412029]
32. Dascal J, Reid M, IsHak WW, Spiegel B, Recacho J, Rosen B, et al. Virtual Reality and Medical Inpatients: A Systematic Review of Randomized, Controlled Trials. Innov Clin Neurosci 2017;14(1-2):14-21 [FREE Full text] [Medline: 28386517]
33. Sato K, Fukumori S, Matsusaki T, Maruo T, Ishikawa S, Nishie H, et al. Nonimmersive virtual reality mirror visual feedback therapy and its application for the treatment of complex regional pain syndrome: an open-label pilot study. Pain Med 2010 Apr 01;11(4):622-629. [doi: 10.1111/j.1526-4637.2010.00819.x] [Medline: 20202141]
34. Shiri S, Feintuch U, Weiss N, Pustilnik A, Geffen T, Kay B, et al. A virtual reality system combined with biofeedback for treating pediatric chronic headache--a pilot study. Pain Med 2013 May 01;14(5):621-627. [doi: 10.1111/pme.12083] [Medline: 23659372]
35. Garcia-Palacios A, Herrero R, Vizcaíno Y, Belmonte MA, Castilla D, Molinari G, et al. Integrating Virtual Reality With Activity Management for the Treatment of Fibromyalgia. The Clinical Journal of Pain 2015;31(6):564-572. [doi: 10.1097/ajp.0000000000000196]
36. Herrero R, García-Palacios A, Castilla D, Molinari G, Botella C. Virtual reality for the induction of positive emotions in the treatment of fibromyalgia: a pilot study over acceptability, satisfaction, and the effect of virtual reality on mood. Cyberpsychol Behav Soc Netw 2014 Jun;17(6):379-384. [doi: 10.1089/cyber.2014.0052] [Medline: 24892201]
37. Sulea C, Soomro A, Boyd C, Wiederhold BK. Pain management in virtual reality: a comprehensive research chart. Cyberpsychol Behav Soc Netw 2014 Jun;17(6):402-413. [doi: 10.1089/cyber.2014.1514] [Medline: 24892205]
38. Wiederhold BK, Gao K, Sulea C, Wiederhold MD. Virtual reality as a distraction technique in chronic pain patients. Cyberpsychol Behav Soc Netw 2014 Jun;17(6):346-352 [FREE Full text] [doi: 10.1089/cyber.2014.0207] [Medline: 24892196]
39. Ahern M, Dean LV, Stoddard CC, Agrawal A, Kim K, Cook CE, et al. The Effectiveness of Virtual Reality in Patients With Spinal Pain: A Systematic Review and Meta-Analysis. Pain Pract 2020 Jul;20(6):656-675. [doi: 10.1111/papr.12885] [Medline: 32196892]
40. Gumaa M, Rehan Youssef A. Is Virtual Reality Effective in Orthopedic Rehabilitation? A Systematic Review and Meta-Analysis. Phys Ther 2019 Oct 28;99(10):1304-1325. [doi: 10.1093/ptj/pzz093] [Medline: 31343702]
41. Nambi G, Abdelbasset WK, Alrawaili SM, Alsubaie SF, Abodonya AM, Saleh AK. Virtual reality or isokinetic training; its effect on pain, kinesiophobia and serum stress hormones in chronic low back pain: A randomized controlled trial. THC 2021 Jan 15;29(1):155-166. [doi: 10.3233/thc-202301]
42. Nambi G, Abdelbasset WK, Elsayed SH, Alrawaili SM, Abodonya AM, Saleh AK, et al. Comparative Effects of Isokinetic Training and Virtual Reality Training on Sports Performances in University Football Players with Chronic Low Back Pain-Randomized Controlled Study. Evid Based Complement Alternat Med 2020 Jun 16;2020:2981273-2981210 [FREE Full text] [doi: 10.1155/2020/2981273] [Medline: 32617104]
43. Garcia LM, Darnall BD, Krishnamurthy P, Mackey IM, Sackman J, Louis R, et al. Self-Administered Behavioral Skills-Based At-Home Virtual Reality Therapy for Chronic Low Back Pain: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2021 Jan 19;10(1):e25291 [FREE Full text] [doi: 10.2196/25291] [Medline: 33464215]
44. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2. Medical Care 2003;41(11):1284-1292. [doi: 10.1097/01.mlr.0000093487.78664.3c]
45. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res 2005 Feb;58(2):163-171. [doi: 10.1016/j.jpsychores.2004.09.006] [Medline: 15820844]
46. Dworkin R, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
47. Dworkin R, Turk DC, McDermott MP, Peirce-Sandner S, Burke LB, Cowan P, et al. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. Pain 2009 Dec;146(3):238-244. [doi: 10.1016/j.pain.2009.08.019] [Medline: 19836888]

48. Turk DC, Dworkin RH, Burke LB, Gershon R, Rothman M, Scott J, et al. Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. Pain 2006 Dec 05;125(3):208-215. [doi: 10.1016/j.pain.2006.09.028] [Medline: 17069973]
49. Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D, Carragee E, et al. Report of the NIH Task Force on research standards for chronic low back pain. Pain Med 2014 Aug 01;15(8):1249-1267. [doi: 10.1111/pme.12538] [Medline: 25132307]
50. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Med 2010 Mar 24;7(3):e1000251 [FREE Full text] [doi: 10.1371/journal.pmed.1000251] [Medline: 20352064]
51. Grant S, CONSORT-SPI Group. The CONSORT-SPI 2018 extension: a new guideline for reporting social and psychological intervention trials. Addiction 2019 Jan 30;114(1):4-8. [doi: 10.1111/add.14411] [Medline: 30091280]
52. US Department of Agriculture. Rural-Urban Commuting Area Codes. URL: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/ [accessed 2020-11-01]
53. Buckenmaier CC, Galloway KT, Polomano RC, McDuffie M, Kwon N, Gallagher RM. Preliminary validation of the Defense and Veterans Pain Rating Scale (DVPRS) in a military population. Pain Med 2013 Jan 01;14(1):110-123. [doi: 10.1111/j.1526-4637.2012.01516.x] [Medline: 23137169]
54. Nassif TH, Hull A, Holliday SB, Sullivan P, Sandbrink F. Concurrent Validity of the Defense and Veterans Pain Rating Scale in VA Outpatients. Pain Med 2015 Nov 01;16(11):2152-2161. [doi: 10.1111/pme.12866] [Medline: 26257151]
55. Amtmann D, Cook KF, Jensen MP, Chen WH, Choi S, Revicki D, et al. Development of a PROMIS item bank to measure pain interference. Pain 2010 Jul;150(1):173-182 [FREE Full text] [doi: 10.1016/j.pain.2010.04.025] [Medline: 20554116]
56. Fries JF, Krishnan E, Rose M, Lingala B, Bruce B. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. Arthritis Res Ther 2011;13(5):R147 [FREE Full text] [doi: 10.1186/ar3461] [Medline: 21914216]
57. Yu L, Buysse DJ, Germain A, Moul DE, Stover A, Dodds NE, et al. Development of short forms from the PROMIS™ sleep disturbance and Sleep-Related Impairment item banks. Behav Sleep Med 2011 Dec 28;10(1):6-24 [FREE Full text] [doi: 10.1080/15402002.2012.636266] [Medline: 22250775]
58. Amtmann D, Cook KF, Johnson KL, Cella D. The PROMIS initiative: involvement of rehabilitation stakeholders in development and examples of applications in rehabilitation research. Arch Phys Med Rehabil 2011 Oct;92(10 Suppl):S12-S19 [FREE Full text] [doi: 10.1016/j.apmr.2011.04.025] [Medline: 21958918]
59. The Assessment Center. URL: https://www.assessmentcenter.net/ [accessed 2021-02-13]
60. Gershon RC, Rothrock N, Hanrahan R, Bass M, Cella D. The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. J Appl Meas 2010;11(3):304-314 [FREE Full text] [Medline: 20847477]
61. Varni JW, Magnus B, Stucky BD, Liu Y, Quinn H, Thissen D, et al. Psychometric properties of the PROMIS ® pediatric scales: precision, stability, and comparison of different scoring and administration options. Qual Life Res 2014 May 2;23(4):1233-1243 [FREE Full text] [doi: 10.1007/s11136-013-0544-0] [Medline: 24085345]
62. Fischer HF, Rose M. Scoring Depression on a Common Metric: A Comparison of EAP Estimation, Plausible Value Imputation, and Full Bayesian IRT Modeling. Multivariate Behav Res 2019 Sep 20;54(1):85-99. [doi: 10.1080/00273171.2018.1491381] [Medline: 30235003]
63. Hung M, Baumhauer JF, Latt LD, Saltzman CL, SooHoo NF, Hunt KJ. Validation of PROMIS ® Physical Function computerized adaptive tests for orthopaedic foot and ankle outcome research. Clin Orthop Relat Res 2013 Nov;471(11):3466-3474 [FREE Full text] [doi: 10.1007/s11999-013-3097-1] [Medline: 23749433]
64. Sturgeon JA, Darnall BD, Kao MJ, Mackey SC. Physical and psychological correlates of fatigue and physical function: a Collaborative Health Outcomes Information Registry (CHOIR) study. J Pain 2015 Mar;16(3):291-8.e1 [FREE Full text] [doi: 10.1016/j.jpain.2014.12.004] [Medline: 25536536]
65. Sturgeon J, Dixon EA, Darnall BD, Mackey SC. Contributions of physical function and satisfaction with social roles to emotional distress in chronic pain: a Collaborative Health Outcomes Information Registry (CHOIR) study. Pain 2015 Dec;156(12):2627-2633 [FREE Full text] [doi: 10.1097/j.pain.0000000000000313] [Medline: 26230739]
66. Sullivan M, Bishop S, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychological Assessment 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
67. Darnall BD, Sturgeon JA, Cook KF, Taub CJ, Roy A, Burns JW, et al. Development and Validation of a Daily Pain Catastrophizing Scale. J Pain 2017 Sep;18(9):1139-1149 [FREE Full text] [doi: 10.1016/j.jpain.2017.05.003] [Medline: 28528981]
68. Nicholas MK, McGuire BE, Asghari A. A 2-item short form of the Pain Self-efficacy Questionnaire: development and psychometric evaluation of PSEQ-2. J Pain 2015 Feb;16(2):153-163. [doi: 10.1016/j.jpain.2014.11.002] [Medline: 25463701]
69. Fish R, McGuire B, Hogan M, Morrison TG, Stewart I. Validation of the chronic pain acceptance questionnaire (CPAQ) in an Internet sample and development and preliminary validation of the CPAQ-8. Pain 2010 Jun;149(3):435-443. [doi: 10.1016/j.pain.2009.12.016] [Medline: 20188472]
70. Brooke J. SUS: A 'Quick and Dirty' Usability Scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation In Industry. London, UK: Taylor and Francis; 1996:189-194.

71. Opioid Oral Morphine Milligram Equivalent (MME) Conversion Factors. 2017. URL: https://www.cms.gov/Medicare/Prescription-Drug-coverage/PrescriptionDrugCovContra/Downloads/Opioid-Morphine-EQ-Conversion-Factors-Aug-2017.pdf [accessed 2021-02-03]
72. Rosenthal R, Rosnow RL. Essentials of Behavioral Research: Methods and Data Analysis (3rd edition). New York, NY: McGraw-Hill; 2008.
73. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
74. Geraghty AWA, Torres LD, Leykin Y, Pérez-Stable EJ, Muñoz RF. Understanding attrition from international Internet health interventions: a step towards global eHealth. Health Promot Int 2013 Sep;28(3):442-452 [FREE Full text] [doi: 10.1093/heapro/das029] [Medline: 22786673]
75. Geraghty AWA, Wood AM, Hyland ME. Attrition from self-directed interventions: investigating the relationship between psychological predictors, intervention content and dropout from a body dissatisfaction intervention. Soc Sci Med 2010 Jul;71(1):30-37. [doi: 10.1016/j.socscimed.2010.03.007] [Medline: 20400220]
76. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
77. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for Methodology of Virtual Reality Clinical Trials in Health Care by an International Working Group: Iterative Study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]
78. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14. [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
79. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
80. Saredakis D, Szpak A, Birckhead B, Keage HAD, Rizzo A, Loetscher T. Factors Associated With Virtual Reality Sickness in Head-Mounted Displays: A Systematic Review and Meta-Analysis. Front Hum Neurosci 2020;14:96 [FREE Full text] [doi: 10.3389/fnhum.2020.00096] [Medline: 32300295]

Abbreviations

CBT: cognitive behavioral therapy
cLBP: chronic low back pain
CPAQ-8: 8-item Chronic Pain Acceptance Questionnaire
DVPRS: Defense and Veterans Pain Rating Scale
MME: morphine milligram equivalent
OTC: over the counter
PCS: Pain Catastrophizing Scale
PGIC: Patient's Global Impression of Change
PROMIS: Physical Function and Sleep Disturbance
PSEQ-2: 2-item Pain Self-Efficacy Questionnaire
RCT: randomized controlled trial
RUCA: rural–urban commuting area
SUS: System Usability Scale
VR: virtual reality

Edited by G Eysenbach; submitted 23.12.20; peer-reviewed by M Linnett, T Jones, S Kardes, P Gazerani; comments to author 06.01.21; revised version received 11.01.21; accepted 13.01.21; published 22.02.21.

Please cite as:
Garcia LM, Birckhead BJ, Krishnamurthy P, Sackman J, Mackey IG, Louis RG, Salmasi V, Maddox T, Darnall BD
An 8-Week Self-Administered At-Home Behavioral Skills-Based Virtual Reality Program for Chronic Low Back Pain: Double-Blind, Randomized, Placebo-Controlled Trial Conducted During COVID-19
J Med Internet Res 2021;23(2):e26292
URL: https://www.jmir.org/2021/2/e26292
doi:10.2196/26292
PMID:33484240

Review

Measurement of Digital Literacy Among Older Adults: Systematic Review

Sarah Soyeon Oh[1], PhD; Kyoung-A Kim[2], PhD; Minsu Kim[3]; Jaeuk Oh[3]; Sang Hui Chu[1], PhD; JiYeon Choi[1], PhD

[1]Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University, Seoul, Republic of Korea

[2]Department of Nursing, Yeoju Institute of Technology, Yeoju, Gyeonggi-do, Republic of Korea

[3]College of Nursing, Yonsei University, Seoul, Republic of Korea

Corresponding Author:
JiYeon Choi, PhD
Mo-Im Kim Nursing Research Institute, College of Nursing, Yonsei University
50 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 3301
Email: jychoi610@yuhs.ac

Related Articles:

This is a corrected version. See correction statement: https://www.jmir.org/2021/3/e28211/

This is a corrected version. See correction statement: https://www.jmir.org/2021/6/e30828

Abstract

Background: Numerous instruments are designed to measure digital literacy among the general population. However, few studies have assessed the use and appropriateness of these measurements for older populations.

Objective: This systematic review aims to identify and critically appraise studies assessing digital literacy among older adults and to evaluate how digital literacy instruments used in existing studies address the elements of age-appropriate digital literacy using the European Commission's Digital Competence (DigComp) Framework.

Methods: Electronic databases were searched for studies using validated instruments to assess digital literacy among older adults. The quality of all included studies was evaluated using the Crowe Critical Appraisal Tool (CCAT). Instruments were assessed according to their ability to incorporate the competence areas of digital literacy as defined by the DigComp Framework: (1) information and data literacy, (2) communication and collaboration, (3) digital content creation, (4) safety, and (5) problem-solving ability, or attitudes toward information and communication technology use.

Results: Searches yielded 1561 studies, of which 27 studies (17 cross-sectional, 2 before and after, 2 randomized controlled trials, 1 longitudinal, and 1 mixed methods) were included in the final analysis. Studies were conducted in the United States (18/27), Germany (3/27), China (1/27), Italy (1/27), Sweden (1/27), Canada (1/27), Iran (1/27), and Bangladesh (1/27). Studies mostly defined older adults as aged ≥50 years (10/27) or ≥60 years (8/27). Overall, the eHealth Literacy Scale (eHEALS) was the most frequently used instrument measuring digital literacy among older adults (16/27, 59%). Scores on the CCAT ranged from 34 (34/40, 85%) to 40 (40/40, 100%). Most instruments measured 1 or 2 of the DigComp Framework's elements, but the Mobile Device Proficiency Questionnaire (MDPQ) measured all 5 elements, including "digital content creation" and "safety."

Conclusions: The current digital literacy assessment instruments targeting older adults have both strengths and weaknesses, relative to their study design, administration method, and ease of use. Certain instrument modalities like the MDPQ are more generalizable and inclusive and thus, favorable for measuring the digital literacy of older adults. More studies focusing on the suitability of such instruments for older populations are warranted, especially for areas like "digital content creation" and "safety" that currently lack assessment. Evidence-based discussions regarding the implications of digitalization for the treatment of older adults and how health care professionals may benefit from this phenomenon are encouraged.

(J Med Internet Res 2021;23(2):e26145) doi:10.2196/26145

KEYWORDS

healthy aging; eHealth; telehealth; mobile health; digital literacy; ehealth literacy; aging; elderly; older adults; review; literacy

Introduction

Background

Adopting digital technology is becoming imperative for all areas of service and business including health care. In the era of global aging, digital technology is viewed as a new opportunity to overcome various challenges associated with aging, such as reduced physical and cognitive function, multiple chronic conditions, and altered social networking [1]. Consistent with this trend, the proportion of older populations using digital technology has increased exponentially [2], although this proportion is still smaller than that of younger generations. According to the latest Digital Economy Outlook Report from the Organization for Economic Cooperation and Development (OECD), 62.8% of 55–74-year-olds are now connected to the internet, as are 96.5% of 16–24-year-olds [3].

Improving the inclusion and engagement of older adults in digital technology is becoming increasingly important for the promotion of their health and function [4]. While numerous studies have measured the digital literacy of younger generations [5,6], few have examined the inclusion of older adults in the research and design of digital technologies. Moreover, existing measures of digital literacy for older adults are generally focused on acceptance models and barriers to adoption [7-9], which fail to consider heterogeneity in user ability. As emphasized by Mannheim et al [10], designs that focus heavily on barriers may be marginalizing older adults by assuming that they are less capable of utilizing digital technologies than their younger counterparts.

For health care professionals, the rapid digitalization of social and health care services has various implications for providing older adults with improved access, knowledge, and behavior [11]. Telehealth platforms are a solution for frailer, older adults to receive medical support remotely [12], while GPS can be used to mine personalized data to locate older patients and track or predict their needs [13]. Internet use is associated with reduced likelihood of depression among the retired, and social networking sites represent an opportunity for older adults to reduce feelings of loneliness through online interactions with family and friends [14]. The increasing number of Alzheimer's disease forums on the microblogging system, Twitter, for example, shows how social networking systems serve as a platform for older individuals to share the latest health-related information with others [13].

Quantifying the digital literacy of older adults is the first step to assist older adults to take advantage of this trend of digitalization in health care. However, when measuring digital literacy among older adults, measures must consider how basic competencies among one age cohort can be harder to achieve for another cohort with fewer information-and-communication-technology experiences and opportunities [15]. In the case of older adults, other age-related factors including life transitions, personal health, attitudes, and economic incentives must also be considered during instrument research and design [16].

Prior Work

To our knowledge, few systematic reviews to date have evaluated instruments of digital literacy for older adults in general, although 1 systematic review of digitally underserved populations attributed poor eHealth literacy to age, as well as language, educational attainment, residential area, and race [17]. Furthermore, the compatibility between these instruments and older adults has not been measured according to a validated framework.

Goal of This Study

Therefore, this systematic review aimed to (1) identify and critically appraise studies that involved the assessment of digital literacy among older adults and (2) evaluate how digital literacy instruments used in existing studies address the elements of age-appropriate digital literacy using the European Commission's Digital Competence (DigComp) Framework [18]. According to DigComp, digital literacy is defined in 5 areas: (1) information and data literacy, (2) communication and collaboration, (3) digital content creation, (4) safety, and (5) problem solving [18]. For this review, we chose the DigComp over other frameworks, such as the International Computer and Information Literacy Study [19] and OECD's Program for the International Assessment of Adult Competencies [20] because the DigComp Framework is the most generalizable across different regions [21] and age groups [15].

Methods

Search Strategy and Data Sources

This systematic review was conducted by searching multiple electronic databases according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [22]. Electronic databases and search engines employed in the initial screening period included PubMed, CINAHL, Embase, and MeSH. The combination of search keywords for each database was summarized in a table (See Multimedia Appendix 1). Keywords were matched to database-specific indexing terms, and searches were not limited to a specific region or study design. However, we limited the year of study to those that were conducted after 2009 for a more recent conceptualization of digital literacy.

The reference lists of identified studies were manually reviewed by a team of academics to prevent relevant studies from being excluded in our search for relevant articles. EndNote X9 was used for database management.

Eligibility Criteria

We included studies that (1) were published in English, (2) targeted older adults, and (3) measured the use of a validated instrument to assess digital literacy. However, publications were excluded if older adults were not the study's main target population. To elaborate, publications targeting the general

population, for example, were excluded from our list of eligible articles as older adults were not the main target population examined.

Exceptions to this rule were studies that compared older populations to younger populations with the aim of addressing the age-related digital divide, like the study by Schneider and colleagues [23] comparing the digital literacy of "baby boomers" (50-65 years old) to that of millennials (18-35 years old).

Study Selection

Using these eligibility criteria, 3 independent investigators (SO, MK, and JO) examined all studies reporting the use of a digital literacy instrument in the databases and search engines. All studies were screened according to their title and excluded if the main target population did not consist of older adults.

Subsequently, abstracts were screened so that non-English studies and studies not assessing digital literacy through a validated instrument could be excluded from our investigation. During this process, any studies that were incapable of providing information on the required general characteristics were excluded.

Last, full-text reviews were performed to ensure that all articles measured the digital literacy of older adults through validated instruments. In this process, investigator-developed questionnaires were included only if authors mentioned that they had been evaluated by experts for face validity. The instruments mentioned in each article were checked to ensure that they were accessible for our quality assessment. All processes were supervised by 2 independent reviewers (SC and JC), and any disagreement was resolved through discussions.

Data Collection

Data on the general characteristics of the included studies included a summary of the year of publication, study design, region where the study was conducted, age of older adults studied, and main literacy instrument used. Regarding the region where the study was conducted, 2 studies were international collaborations, including 1 study between Italy and Sweden [24] and another study between the United States, United Kingdom, and New Zealand [25]. For these 2 studies, the first author's region of study was used in our general characteristics summary.

Quality Assessment

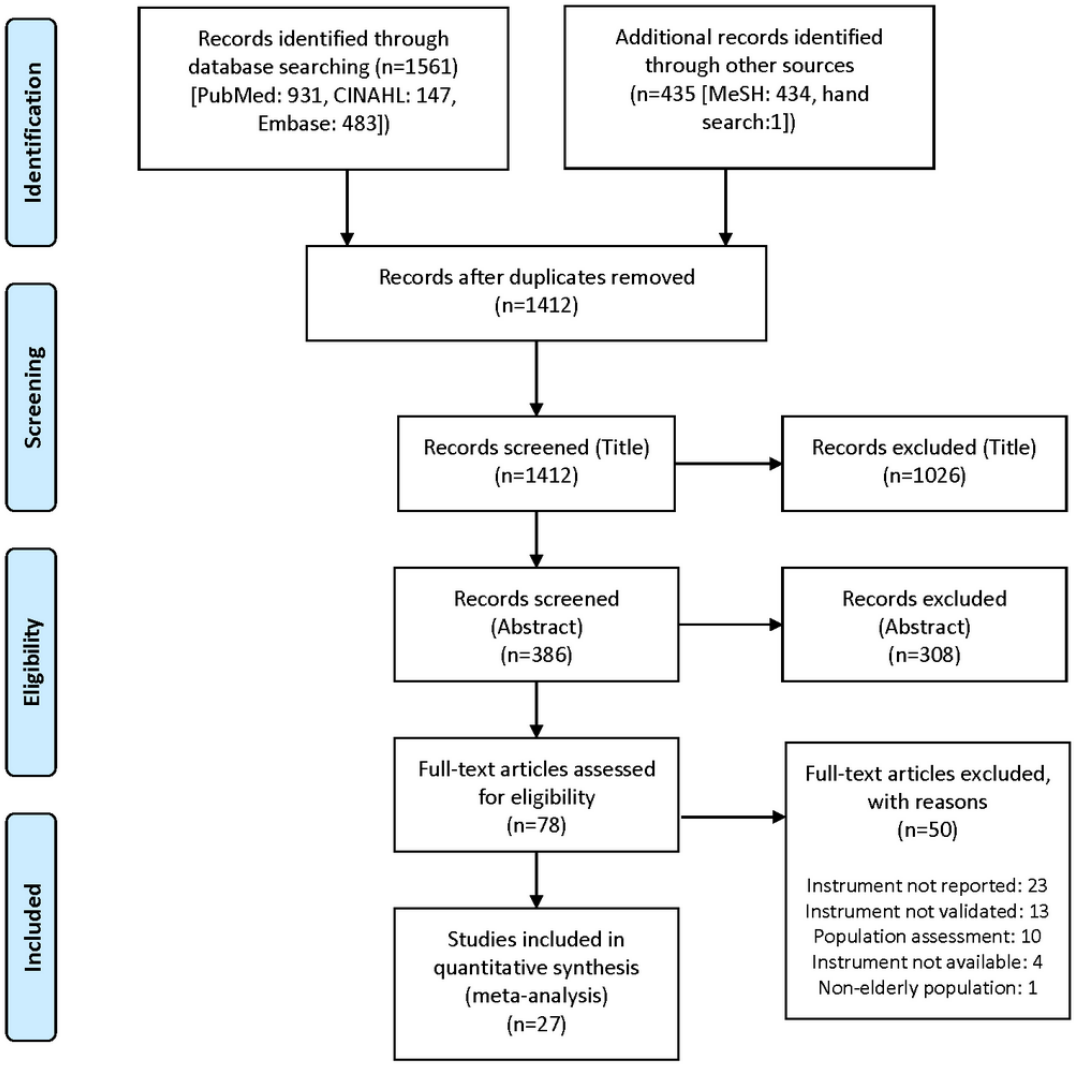
Three independent reviewers (SO, JC, and KK) assessed the quality of each included study using the Crowe Critical Appraisal Tool (CCAT) [26]. The CCAT is a validated quality assessment tool developed to rate research papers in systematic reviews based on a number of criteria relative to research design, variables and analysis, sampling methods, and data collection (Multimedia Appendix 2) [26]. Many systematic reviews targeting older adults have used this tool [27,28] for quality appraisal.

Instruments were also assessed to the DigComp's definition of the 5 areas of digital literacy: (1) information and data literacy (browsing, searching, filtering data), (2) communication and collaboration (interacting, sharing, engaging in citizenship, collaborating), (3) digital content creation (developing, integrating, and re-elaborating digital content; copyright; licenses; programming), (4) safety (protecting devices, protecting personal data and privacy, protecting health and well-being), and (5) problem solving (solving technical problems, identifying needs and technological responses, creatively using digital technologies, identifying digital competence gaps) [18].

Results

The PRISMA flow diagram in Figure 1 summarizes the search results and selection process of all studies included in our synthesis. Overall, the number of records identified in our database was 1561 (PubMed: 931; CINAHL: 147; Embase: 483). The number of additional records identified through other sources was 435 (MeSH: 434, hand search: 1). Of these records, 1412 remained after duplicates were electronically removed. An additional 1026 articles were removed after title screening, and 308 articles were removed after abstract screening.

Figure 1. PRISMA flowchart of the literature search and study selection process.



Study Characteristics

Of the 78 articles assessed for eligibility, 50 were excluded for the following reasons: (1) no report of an instrument for digital literacy despite the title or abstract of the paper alluding to measures of digital literacy (n=23); (2) the instrument presented was not validated (n=13); (3) studies were mainly on population assessments and measured digital literacy only as part of a wider assessment of multiple factors (n=10); (3) instruments were not available in English or in a publicly accessible format (n=4); and (4) the study did not specifically target older adults (n=1). Ultimately, 27 articles were included in our review.

Table 1 provides a general summary of the included studies. While publication years ranged from 2009 to 2020, most articles reviewed were conducted between 2015 and 2020. The majority (17/27, 63%) of included studies were cross-sectional, but 2 studies were pre- and post-test studies, 2 were randomized controlled trials (RCTs), 1 was longitudinal, and 1 was a mixed-method study with both surveys and focus group interviews. Most studies were conducted in the United States (18/27), but some studies were also conducted in Europe (Germany, 3/27; Italy, 1/27; Sweden, 1/27). Studies mostly defined older adults as aged ≥50 years (10/27) or ≥60 years (8/27).

Table 2 presents the detailed characteristics of all 27 included studies. Overall, the eHealth Literacy Scale (eHEALS) [29] was the most frequently used instrument to measure digital literacy among older adults (16/27, 59%). The Unified Theory of Acceptance and Usage of Technology (UTAUT) was also used by 2 studies from Germany [9,30] and 1 study from Bangladesh [31]. Loyd-Gressard's Computer Attitude Scale (CAS) was used in 2 studies that focused heavily on computer anxiety and confidence [32,33]. There was not wide variation in the quality of studies assessed via the CCAT, with scores ranging from 34 (34/40, 85%) to 40 (40/40, 100%) of a total of 40 points.

Table 1. Summary of included studies (n=27).

Categories	Results, n (%)
Year of study publication	
2009-2010	2 (7)
2011-2012	3 (11)
2013-2014	2 (7)
2015-2016	6 (22)
2017-2018	9 (33)
2019-2020	5 (19)
Study design	
Cross-sectional	17 (63)
Before and after study	2 (7)
Randomized controlled trial	2 (7)
Longitudinal	1 (4)
Mixed methods[a]	1 (4)
Region where study was conducted	
United States	18 (67)
Germany	3 (11)
China	1 (4)
Italy	1 (4)
Sweden	1 (4)
Canada	1 (4)
Iran	1 (4)
Bangladesh	1 (4)
Definition of older adults (years)	
≥50	10 (37)
≥55	4 (15)
≥60	8 (30)
≥65	5 (19)
Main health literacy instrument used	
eHealth Literacy Scale (eHEALS)	16 (59)
Unified Theory of Acceptance and Use of Technology (UTAUT)	3 (11)
Computer Anxiety Scale (CAS)	2 (7)
Technology Acceptance Model (TAM)	2 (7)
Swedish Zimbardo Time Perspective Inventory (S-ZTPI)	1 (4)
Mobile Device Proficiency Questionnaire (MDPQ)	1 (4)
Everyday Technology Use Questionnaire (ETUQ)	1 (4)
Attitudes towards Psychological Online Interventions (APOI)	1 (4)

[a]Survey and focus group interviews.

Table 2. Characteristics of included studies.

Author	Year	Country	Sample size, n	Design	Study aim	Measure	CCAT[a,b] score, points (% of total)
Roque et al [34]	2016	United States	109	Cross-sectional	To validate a new tool for measuring mobile device proficiency across the life span by assessing both basic and advanced proficiencies related to smartphone and tablet use	MDPQ[c], CPQ[d]-12	37 (93)
Zambianchi et al [24]	2019	Italy and Sweden	638	Cross-sectional	To examine the determinants of attitudes towards and use of ICTs[e] in older adults	S-ZTPI[f], ATTQ[g]	38 (95)
Schneider et al [23]	2018	Germany	577	RCT[h]	To examine whether there are any differences in use of an online psychological intervention between generational groups based on Deprexis user data, responses on a questionnaire, and data in the EVIDENT study	APOI[i]	40 (100)
Nagle et al [9]	2012	Germany	52	Cross-sectional	To get a better understanding of the factors affecting older adults' intention towards and usage of computers	UTAUT[j]	34 (85)
Yoon et al [33]	2015	United States	209	Cross-sectional	To examine predictors of computer use and computer anxiety in older Korean Americans	CAS[k]	38 (95)
Cherid et al [35]	2020	Canada	401	Cross-sectional	To identify the current level of technology adoption, health, and eHealth literacy among older adults with a recent fracture, to determine if the use of electronic interventions would be feasible and acceptable in this population	eHEALS[l]	39 (98)
Xie and Bo [36]	2011	United States	146	Cross-sectional	To examine the effects of a theory-driven eHealth literacy intervention for older adults	eHEALS	39 (98)
Tennant et al [37]	2015	United States	393	Cross-sectional	To explore the extent to which sociodemographic, social determinants, and electronic device use influence eHealth literacy and use of Web 2.0 for health information among baby boomers and older adults	eHEALS	36 (90)
Hoogland et al [38]	2020	United States	198	Cross-sectional	To examine age differences in eHealth literacy and use of technology devices/HIT[m] in patients with cancer and characterize receptivity towards using home-based HIT to communicate with the oncology care team	eHEALS	36 (90)
Price-Haywood et al [39]	2017	United States	247	Cross-sectional	To examine relationships between portal usages, interest in health-tracking tools, and eHealth literacy and to solicit practical solutions to encourage technology adoption.	eHEALS	37 (93)
Paige et al [40]	2018	United States	830	Cross-sectional	To examine the structure of eHEALS scores and the degree of measurement invariance among US adults representing the following generations: millennials (18-35 years old), Generation X (36-51 years old), baby boomers (52-70 years old), and the silent generation (71-84 years old)	eHEALS	38 (95)
Aponte et al [41]	2017	United States	20	Cross-sectional	To explore the experiences of older Hispanics with type 2 diabetes in using the internet for diabetes management	eHEALS	37 (93)
Xie and Bo [42]	2011	United States	124	Cross-sectional	To generate scientific knowledge about the potential impact of learning methods and information presentation channels on older adults' eHealth literacy	eHEALS	39 (98)

Author	Year	Country	Sample size, n	Design	Study aim	Measure	CCAT[a,b] score, points (% of total)
Sudbury-Riley and Lynn [25]	2017	United States, United Kingdom, New Zealand	996	Cross-sectional	To examine the factorial validity and measurement invariance of the eHEALS among baby boomers in the United States, the United Kingdom, and New Zealand who had used the internet to search for health information in the last 6 months	eHEALS	35 (88)
Noblin et al [43]	2017	United States	181	Cross-sectional	To determine the willingness of older adults to use health information from a variety of sources	eHEALS	38 (95)
Cajita et al [44]	2018	United States	129	Cross-sectional	To examine factors that influence intention to use mobile technology in health care (mHealth) among older adults with heart failure	TAM[n]	37 (93)
Lin et al [45]	2019	Iran	468	Longitudinal	To examine the temporal associations between eHealth literacy, insomnia, psychological distress, medication adherence, quality of life, and cardiac events among older patients with heart failure	eHEALS	39 (98)
Chu et al [32]	2009	United States	137	RCT	To measure the psychosocial influences of computer anxiety, computer confidence, and computer self-efficacy in older adults at 6 meal congregate sites	CAS	40 (100)
Rosenberg et al [46]	2009	Sweden	157	Cross-sectional	To measure the perceived difficulty in everyday technology use such as remote controls, cell phones, and microwave ovens by older adults with or without cognitive deficits	ETUQ[o]	37 (93)
Stellefson et al [47]	2017	United States	283	Cross-sectional	To examine the reliability and internal structure of eHEALS data collected from older adults aged ≥50 years responding to items over the telephone	eHEALS	36 (90)
Chung et al [48]	2015	United States	866	Cross-sectional	To test the psychometric aspects of the eHEALS for older adults using secondary data analysis	eHEALS	36 (90)
Li et al [49]	2020	China	1201	Cross-sectional	To examine the associations among health-promoting lifestyles, eHealth literacy, and cognitive health in older adults	eHEALS	37 (93)
Choi et al [29]	2013	United States	980	Mixed methods	To examine internet use patterns, reasons for discontinued use, eHealth literacy, and attitudes toward computer or internet use among low-income homebound individuals aged ≥60 years in comparison to their younger counterparts (homebound adults <60 years old)	eHEALS, ATC/IQ[p]	37 (93)
Moore et al [8]	2015	United States	30	Cross-sectional	To offer design considerations in developing internet-based hearing health care for older adults by analyzing and discussing the relationship between chronological age, computer skills, and the acceptance of internet-based hearing health care	TAM	35 (88)
Hoque et al [31]	2017	Bangladesh	300	Cross-sectional	To develop a theoretical model based on the UTAUT and then empirically test it to determine the key factors influencing elderly users' intention to adopt and use mHealth services	UTAUT	36 (90)
Niehaves et al [30]	2014	Germany	150	Cross-sectional	To study the intentions of the elderly with regard to internet use and identify important influencing factors	UTAUT	35 (88)

Author	Year	Country	Sample size, n	Design	Study aim	Measure	CCAT[a,b] score, points (% of total)
Aponte et al [41]	2017	United States	100	Cross-sectional	To examine the validity of the Spanish version of the eHEALS with an older Hispanic population from a number of Spanish-language countries living in New York City	eHEALS	37 (93)

[a]CCAT: Crowe Critical Appraisal Tool.

[b]Total CCAT score is 40 points.

[c]MDPQ: Mobile Device Proficiency Questionnaire.

[d]CPQ: Computer Proficiency Questionnaire.

[e]ICTs: information and communication technologies.

[f]S-ZTPI: Swedish Zimbardo Time Perspective Inventory.

[g]ATTQ: Attitudes Toward Technologies Questionnaire.

[h]RCT: randomized controlled trial.

[i]APOI: Attitudes towards Psychological Online Interventions.

[j]UTAUT: Unified Theory of Acceptance and Use of Technology.

[k]CAS: Computer Attitude Scale.

[l]eHEALS: eHealth Literacy Scale.

[m]HIT: health information technology.

[n]TAM: Adapted Technology Acceptance Model.

[o]ETUQ: Everyday Technology Use Questionnaire.

[p]ATC/IQ: Attitudes Toward Computer/Internet Questionnaire.

As seen in Table 3, all instruments were analyzed for quality assessment to assess which DigComp elements of digital literacy were met [18]. Studies mostly satisfied 1 or 2 aspects of the information and data literacy criteria, but the Mobile Device Proficiency Questionnaire (MDPQ) satisfied all 5 elements, including those related to safety and data creation.

Table 3. Inclusion of the European Commission's Digital Competence (DigComp) Framework criteria and quality assessment of the included studies.

Measure	Literacy elements[a]					Mode	Scoring	Reliability, Cronbach α
	1	2	3	4	5			
Attitude Toward Technologies Questionnaire (ATTQ)	O[b]	O	X[c]	X	X	Self-administered	6 5-point Likert questions	0.91 (Italy), 0.92 (Sweden) [50]
Adapted Technology Acceptance Model (TAM)	O	O	X	X	X	Self-administered	6 7-point Likert questions	0.91 (perceived ease of use), 0.97 (perceived usefulness), 0.96 (attitude toward using), 0.70 (actual system use) [51]
Attitudes Toward Computer/Internet Questionnaire (ATC/IQ)	O	X	X	X	X	Interview (semistructured)	10 5-point Likert questions	0.98 (usefulness), 0.94 (ease of use) [52], adapted by Choi and DiNitto [29]
Attitudes Towards Psychological Online Intervention Questionnaire (APOI)	O	X	X	X	X	Self-administered	16 5-point Likert questions	0.77 (total), 0.62 (skepticism and perception of risks), 0.62 (anonymity benefits), 0.64 (technologization threat), 0.72 (confidence in effectiveness) [53]
Computer Attitude Scale (CAS)	X	X	X	O	O	Self-administered	4 10-point Likert questions	0.95 (total), 0.90 (computer anxiety), 0.89 (computer confidence), 0.89 (computer liking), 0.82 (computer usefulness) [54]
eHealth Literacy Scale (eHEALS)	O	O	X	O	O	Self-administered	8 5-point Likert questions	0.88, 0.60-0.84 (range among items) [55]
Computer Proficiency Questionnaire (CPQ)	O	O	O	X	X	Self-administered	33 5-point Likert questions	0.98 (total for CPQ) 0.95 (total for CPQ-12), 0.91 (computer basics), 0.94 (printing), 0.95 (communication), 0.97 (internet), 0.96 (scheduling), 0.86 (multimedia) [56]
Mobile Device Proficiency Questionnaire (MDPQ)	O	O	O	O	O	Self-administered	46 5-point Likert questions	0.75 (MDPQ-46), 0.99 (MDPQ-16) [34]
Unified Theory of Acceptance and Usage of Technology (UTAUT)	O	O	O	X	O	Interview (face-to-face)	15 7-point Likert questions	0.7879-0.9497 [57]

[a]European Commission's Digital Competence (DigComp) Framework criteria of (1) information and data literacy (browsing, searching, filtering data), (2) communication and collaboration (interacting, sharing, engaging in citizenship, collaborating), (3) digital content creation (developing, integrating, and re-elaborating digital content; copyright; licenses; programming), (4) safety (protecting devices, protecting personal data and privacy, protecting health and well-being), and (5) problem solving (solving technical problems, identifying needs and technological responses, creatively using digital technologies, identifying digital competence gaps) [18].

[b]O: included in the questionnaire.

[c]X: not included in the questionnaire.

Discussion

Principal Findings

In this systematic review, we highlighted the importance of digital literacy among older adults and provided a comprehensive overview of the instruments that are being employed to measure their digital literacy. We also illustrated the various strengths and limitations of each instrument, relative to age-appropriateness and suitability for older adults, in accordance with the components of a validated, digital competency framework [18]. Our review is timely because, to the best of our knowledge, few systematic reviews to date have evaluated measurements of digital literacy for older adults specifically.

In the digital era, providing education for patients regarding management of their physical or mental illness or injury, explaining posttreatment home care needs, and managing their diet, nutrition, and exercise are all duties that are beginning to be "digitalized" [58]. Moreover, digital technologies are providing practitioners with more effective and user-centered ways to educate, inform, and treat older patients. For example, in a systematic review of "virtual visits" in home care for older patients, both service users and providers found online visits to be more flexible, easy to arrange, and personal than offline visits [59]. In another study of an internet-based videoconferencing system for frail elderly people in Nordic countries, telehealth was associated with reduced loneliness among 88% of users, while simultaneously reducing the odds of catching a cold during winter months due to leaving the house [60].

Overall, we discovered that while the eHEALS is most frequently used to measure digital literacy among older adults, the MDPQ may be more appropriate for measuring the literacy of older adults. Unlike the eHEALS, the MDPQ attempts to measure older adults' digital content creation capacity (developing, integrating, and re-elaborating digital content; copyright; licenses; programming), which according to the European Commission, can give valuable information regarding an individual's ability to add value to new media for self-expression and knowledge creation [18].

Also, the MDPQ contains numerous items related to data protection and privacy such as "passwords can be created to block/unblock a mobile device" or "search history and

temporary files can be deleted" despite the fact that security was the least measured element of the DigComp Framework among the instruments in our study. Only the CAS, eHEALS, and MDPQ provide items related to data protection and privacy, which is concerning given that older adults comprise a significant proportion of the target population for internet scams or email attacks [61].

In our review of 27 selected articles, more than half (16/27, 59%) used the eHEALS to measure the digital literacy of older adults. Several reasons can be speculated; this instrument is short (8 items), and the questions are simple to understand (eg, "I know how to use the Internet to answer my health questions"). Scholars claim that it is easy to administer to older adults [48]. It should be noted that because of its simplicity, there has been some debate regarding the validity of the eHEALS [62-64]. As described by Jordan and colleagues [64], the eHEALS has a "lack of an explicit definition of the concept that many health literacy indices were developed to measure limit... ...its ability to make fully informed judgments... ...about a person's ability to seek, understand, and use health information."

Studies focusing on similar research aims also employ similar instruments. For example, the CAS was used in 2 studies that focused on computer anxiety and confidence. In the existing body of literature, the CAS has often been used for studies targeting individuals in highly stressful environments such as business graduate students [65], psychiatric inpatients [66], and students studying at a 2-year technical college experiencing "technostress" [67]. As explained by Kelley and Charness [68], older adults "commit more errors in post-training evaluations" than the general population, which may result in greater stress and anxiety. This may demonstrate the suitability of the CAS for older adult populations.

Regarding the overall quality of the included studies evaluated using the CCAT, some variation existed among the studies reviewed. Studies that were cross-sectional or lacked acquisition of written informed consent and used alternate approaches, such as telephone or self-reported, web-based or email surveys, scored poorly in the "design" and "ethical matters" category. Studies also lost marks if there was no flow diagram, there was no mention of design methods in the title of their manuscript, or they had biased sampling methods (convenience sampling, pertaining only to 1 or 2 ethnic groups).

Contrastingly, 2 RCTs in our review received a score of 100% on the CCAT, as they had excellent preliminaries, introductions, study design, sampling methods, data collection methods, ethical matters, results, and discussions. These studies employed performance-based measures like the Attitudes Toward Computer/Internet Questionnaire (ATC-IQ; semistructured interview) and UTAUT model (face-to-face interview), which are more reliable data collection methods than self-administered questionnaires. Performance-based measures like these may be suitable for studies targeting older adults, but it should be noted that clinical environments and personal fitness can greatly influence outcomes, especially if environments contain learners of mixed ability [69], rapid progression [34], and the possibility for embarrassment or discomfort [70]. Positive clinical settings are associated with improved performance, as observed in 1 of the RCTs in our review, where "a combination of patience, perseverance, and peer-to-peer or instructor encouragement, whether with words or a pat on the shoulder" were successful in reducing older adults' stress and anxiety during digital learning [32].

As aforementioned, for older adults, it is important that the research and design of digital technologies encompass the heterogeneity of their capacity. While we believe that instructions should be "clear and understandable" to study participants [34], we also believe that literacy elements that are generalizable to the rest of the population (relative to communication, safety, problem solving, and competence) should be measured for this population as well. As described by Hänninen et al [16], the digital capacity of older adults lies on a continuum or spectrum and can range from actively independent to limited.

Previous studies recommend that, instead of employing the full MDPQ or technology acceptance model (TAM), the shorter, 16-question version [34], or senior version of the TAM (Senior Technology Acceptance & Adoption Model), may be more appropriate for relatively older and frailer populations [7]. User-centeredness in instrument development and measurement is crucial for this population, as the functional status of older adults varies immensely. Furthermore, scales and scoring methods are encouraged to be as inclusive as possible, so that they encompass the diversity in functionality that exists among study subjects.

Limitations

Ultimately, many limitations exist in our review. First, it is important to mention that the association between age and digital capacity is controversial among certain scholars who argue that age-based divisions are too simplistic [23] and unclear [71] to explain the digital divide. In the Netherlands, for example, "digital natives" do not appear to exist, and other factors like life stages and socialization are considered to be more relevant proxies of digital literacy than age [71]. Also, in a German study, perceptions of threat due to technologization were perceived as the main predictors of digital capacity, rather than age itself [23]. Older adults with lower perceptions of threat could be digitally fluent, just as younger adults with higher perceptions of threat could be digitally illiterate. Future questionnaires should consider measuring this factor in depth and the possible interaction that it has with age in predicting digital capacity outcomes.

Likewise, digital literacy is a process-oriented skill, and measuring it in isolation may be inaccurate for quantifying an individual's skillset [72]. In the Lily Model, Norman and Skinner [72] posit that there are 6 core skills of literacy: traditional, media, information, computer, scientific, and health. Not only are these skills heavily interconnected with one another but also only an in-depth analysis of all 6 can fully contextualize an individual's personal, social, and environmental contexts [72]. For example, computer literacy may be heavily influenced by an individual's ability to understand and read passages (traditional literacy) as well as their ability to find information on a topic (information literacy) and understand certain scientific terms (science literacy). Because these literacy types are

interconnected, only an in-depth analysis of all 6 may accurately measure an individual's knowledge.

Also, as observed in our review, many of the investigated instruments, including the Attitudes Toward Technologies Questionnaire, TAM, ATC-IQ, APOI, and CAS, measured attitudes or perceptions toward technology rather than digital aptitude itself. While studies on attitude are important, the lack of measures examining older adults' abilities to use information and communications technology was an unexpected limitation of the reviews studied.

Last, even though previous studies have argued that the DigComp Framework is one of the broadest and most generalizable frameworks for assessing digital literacy measures [15,21], it is undeniable that certain types of survey error are more likely to occur among older populations relative to memory loss, health problems, sensory and cognitive impairments, and personal or motivational factors that influence their ability to participate in an investigation [73]. The author and editors of this framework specifically mention in their proposal that, because they adopted a "general" rather than "individual" approach, their framework should be considered only as a starting point in interpreting digital competence among different age groups [18].

Conclusions

In conclusion, more studies are required so that the measurement of digital literacy among older adults can become more elaborate and specific. Digital literacy evidently has strong associations with the utility of information and communications technologies that promote physical and mental well-being among older adults. Further assessments and studies of digital literacy among older adults that overcome the limitations of existing research and measurement designs would allow for better allocation of support and resources to address the diverse health care needs of this growing but vulnerable population.

Acknowledgments

This research was supported by the Brain Korea 21 FOUR Project funded by the National Research Foundation of Korea, Yonsei University College of Nursing. This study received funding from the National Research Foundation of Korea (grant number 2020R1A6A1A0304198911 [SO, KK, and JC]), the Ministry of Education of the Republic of Korea, and Yonsei University College of Nursing Faculty Research Fund (6-2020-0188 [SHC and JC]).

Authors' Contributions

JC and SC conceptualized and supervised the study. JC, SC, and SO developed the methodology. SO, MK, and JO screened the studies and performed the formal analysis. KK and JC performed the validation. SO, KK, SC, and JC wrote, reviewed, and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of database (DB) search terms.
[DOCX File , 13 KB - jmir_v23i2e26145_app1.docx]

Multimedia Appendix 2
Crowe Critical Appraisal Tool (CCAT) form.
[DOCX File , 571 KB - jmir_v23i2e26145_app2.docx]

References

1. Maharani A, Pendleton N, Leroi I. Hearing Impairment, Loneliness, Social Isolation, and Cognitive Function: Longitudinal Analysis Using English Longitudinal Study on Ageing. Am J Geriatr Psychiatry 2019 Dec;27(12):1348-1356. [doi: 10.1016/j.jagp.2019.07.010] [Medline: 31402088]
2. Anderson M, Perrin A. Tech adoption climbs among older adults. Pew Research Center. 2017 May 17. URL: https://www.pewresearch.org/internet/2017/05/17/tech-adoption-climbs-among-older-adults/ [accessed 2021-01-23]
3. OECD Secretariat. OECD Digital Economy Outlook 2017. Organisation for Economic Cooperation and Development (OECD). 2017 Oct 11. URL: https://www.oecd.org/digital/oecd-digital-economy-outlook-2017-9789264276284-en.htm [accessed 2021-01-23]
4. Rogers SE. Bridging the 21st Century Digital Divide. TechTrends 2016 Mar 30;60(3):197-199. [doi: 10.1007/s11528-016-0057-0]
5. Akçayır M, Dündar H, Akçayır G. What makes you a digital native? Is it enough to be born after 1980? Computers in Human Behavior 2016 Jul;60:435-440. [doi: 10.1016/j.chb.2016.02.089]
6. Oblinger DG, Oblinger JL, Lippincott JK, editors. Educating the net generation. Boulder, CO: EDUCAUSE; 2005.

7. Chen K, Chan AHS. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014 Mar 24;57(5):635-652. [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]
8. Moore AN, Rothpletz AM, Preminger JE. The Effect of Chronological Age on the Acceptance of Internet-Based Hearing Health Care. Am J Audiol 2015 Sep;24(3):280-283. [doi: 10.1044/2015_aja-14-0082]
9. Nägle S, Schmidt L. Computer acceptance of older adults. Work 2012;41 Suppl 1:3541-3548. [doi: 10.3233/WOR-2012-0633-3541]
10. Mannheim I, Schwartz E, Xi W, Buttigieg SC, McDonnell-Naughton M, Wouters EJM, et al. Inclusion of Older Adults in the Research and Design of Digital Technology. Int J Environ Res Public Health 2019 Oct 02;16(19):3718 [FREE Full text] [doi: 10.3390/ijerph16193718] [Medline: 31581632]
11. DeWalt DA, Broucksou KA, Hawk V, Brach C, Hink A, Rudd R, et al. Developing and testing the health literacy universal precautions toolkit. Nurs Outlook 2011 Mar;59(2):85-94 [FREE Full text] [doi: 10.1016/j.outlook.2010.12.002]
12. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(7):1147-1148 [FREE Full text] [doi: 10.1093/jamia/ocaa078]
13. Cotten S. Examining the Roles of Technology in Aging and Quality of Life. J Gerontol B Psychol Sci Soc Sci 2017 Sep 01;72(5):823-826. [doi: 10.1093/geronb/gbx109]
14. McCausland L, Falk NL. From dinner table to digital tablet: technology's potential for reducing loneliness in older adults. J Psychosoc Nurs Ment Health Serv 2012 May 25;50(5):22-26 [FREE Full text] [Medline: 22533840]
15. Jin K, Reichert F, Cagasan LP, de la Torre J, Law N. Measuring digital literacy across three age cohorts: Exploring test dimensionality and performance differences. Computers & Education 2020 Nov;157:103968. [doi: 10.1016/j.compedu.2020.103968]
16. Hänninen R, Taipale S, Luostari R. Exploring heterogeneous ICT use among older adults:The warm experts' perspective. New Media & Society 2020 May 06:146144482091735. [doi: 10.1177/1461444820917353]
17. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: A systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016 Feb 24;41(1):1-19. [doi: 10.3109/17538157.2014.948171]
18. Ferrari A, Punie Y, Brecko B. DIGCOMP: A framework for developing and understanding digital competence in Europe. Luxembourg: Publications Office of the European Union; 2013.
19. Fraillon J, Ainley J. The IEA international study of computer and information literacy (ICILS). Camberwell, Victoria: Australian Council for Educational Research; 2013. URL: https://cms.education.gov.il/NR/rdonlyres/A07C37A9-8709-4B9B-85C0-63B9D57F1715/167512/ICILSProjectDescription1.pdf [accessed 2021-01-23]
20. Literacy, Numeracy and Problem Solving in Technology-Rich Environments: Framework for the OECD Survey of Adult Skills. Organisation for Economic Cooperation and Development (OECD). Paris, France: OECD Publishing; 2012. URL: https://www.oecd-ilibrary.org/education/literacy-numeracy-and-problem-solving-in-technology-rich-environments_9789264128859-en [accessed 2021-01-15]
21. Law N, Woo D, de Ia J, Wong G. A global framework of reference on digital literacy skills for indicator 4.4.2. UNESCO Institute for Statistics. 2018. URL: https://unesdoc.unesco.org/ark:/48223/pf0000265403 [accessed 2021-01-23]
22. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, the PRISMA-DTA Group, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA 2018 Jan 23;319(4):388-396. [doi: 10.1001/jama.2017.19163]
23. Schneider BC, Schröder J, Berger T, Hohagen F, Meyer B, Späth C, et al. Bridging the "digital divide": A comparison of use and effectiveness of an online intervention for depression between Baby Boomers and Millennials. J Affect Disord 2018 Aug 15;236:243-251. [doi: 10.1016/j.jad.2018.04.101] [Medline: 29751239]
24. Zambianchi M, Rönnlund M, Carelli MG. Attitudes Towards and Use of Information and Communication Technologies (ICTs) Among Older Adults in Italy and Sweden: the Influence of Cultural Context, Socio-Demographic Factors, and Time Perspective. J Cross Cult Gerontol 2019 Sep 11;34(3):291-306. [doi: 10.1007/s10823-019-09370-y] [Medline: 31079248]
25. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the Measurement Properties of the eHealth Literacy Scale (eHEALS) Among Baby Boomers: A Multinational Test of Measurement Invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
26. Crowe M. Crowe critical appraisal tool (CCAT) user guide. Conchra House. 2013. URL: https://conchra.com.au/wp-content/uploads/2015/12/CCAT-user-guide-v1.4.pdf [accessed 2021-01-23]
27. Donnelly N, Hickey A, Burns A, Murphy P, Doyle F. Systematic review and meta-analysis of the impact of carer stress on subsequent institutionalisation of community-dwelling older people. PLoS One 2015 Jun 2;10(6):e0128213 [FREE Full text] [doi: 10.1371/journal.pone.0128213] [Medline: 26035433]
28. Sezgin D, Hendry A, Liew A, O'Donovan M, Salem M, Carriazo AM, et al. Transitional palliative care interventions for older adults with advanced non-malignant diseases and frailty: a systematic review. JICA 2020 Jun 17;28(4):387-403. [doi: 10.1108/jica-02-2020-0012]
29. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]

30. Niehaves B, Plattfaut R. Internet adoption by the elderly: employing IS technology acceptance theories for understanding the age-related digital divide. European Journal of Information Systems 2017 Dec 19;23(6):708-726. [doi: 10.1057/ejis.2013.19]
31. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: An extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
32. Chu AYM. Psychosocial influences of computer anxiety, computer confidence, and computer self-efficacy with online health information in older adults. Denton, TX: Texas Woman's University; 2008.
33. Yoon H, Jang Y, Xie B. Computer Use and Computer Anxiety in Older Korean Americans. J Appl Gerontol 2016 Sep 09;35(9):1000-1010. [doi: 10.1177/0733464815570665] [Medline: 25698717]
34. Roque NA, Boot WR. A New Tool for Assessing Mobile Device Proficiency in Older Adults: The Mobile Device Proficiency Questionnaire. J Appl Gerontol 2018 Feb 11;37(2):131-156. [doi: 10.1177/0733464816642582] [Medline: 27255686]
35. Cherid C, Baghdadli A, Wall M, Mayo NE, Berry G, Harvey EJ, et al. Current level of technology use, health and eHealth literacy in older Canadians with a recent fracture-a survey in orthopedic clinics. Osteoporos Int 2020 Jul 28;31(7):1333-1340. [doi: 10.1007/s00198-020-05359-3] [Medline: 32112118]
36. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011 Nov 03;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]
37. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
38. Hoogland AI, Mansfield J, Lafranchise EA, Bulls HW, Johnstone PA, Jim HS. eHealth literacy in older adults with cancer. J Geriatr Oncol 2020 Jul;11(6):1020-1022. [doi: 10.1016/j.jgo.2019.12.015] [Medline: 31917114]
39. Price-Haywood EG, Harden-Barrios J, Ulep R, Luo Q. eHealth Literacy: Patient Engagement in Identifying Strategies to Encourage Use of Patient Portals Among Older Adults. Popul Health Manag 2017 Dec;20(6):486-494. [doi: 10.1089/pop.2016.0164] [Medline: 28384076]
40. Paige SR, Miller MD, Krieger JL, Stellefson M, Cheong J. Electronic Health Literacy Across the Lifespan: Measurement Invariance Study. J Med Internet Res 2018 Jul 09;20(7):e10434 [FREE Full text] [doi: 10.2196/10434] [Medline: 29986848]
41. Aponte J, Nokes K. Electronic health literacy of older Hispanics with diabetes. Health Promot Int 2017 Jun 01;32(3):482-489. [doi: 10.1093/heapro/dav112] [Medline: 26681770]
42. Xie B. Experimenting on the impact of learning methods and information presentation channels on older adults' e-health literacy. J. Am. Soc. Inf. Sci 2011 Jun 06;62(9):1797-1807. [doi: 10.1002/asi.21575]
43. Noblin AM, Rutherford A. Impact of Health Literacy on Senior Citizen Engagement in Health Care IT Usage. Gerontol Geriatr Med 2017 Apr 26;3:2333721417706300 [FREE Full text] [doi: 10.1177/2333721417706300] [Medline: 28508022]
44. Cajita MI, Hodgson NA, Budhathoki C, Han H. Intention to Use mHealth in Older Adults With Heart Failure. The Journal of Cardiovascular Nursing 2017;32(6):E1-E7. [doi: 10.1097/jcn.0000000000000401]
45. Lin C, Ganji M, Griffiths MD, Bravell ME, Broström A, Pakpour AH. Mediated effects of insomnia, psychological distress and medication adherence in the association of eHealth literacy and cardiac events among Iranian older patients with heart failure: a longitudinal study. Eur J Cardiovasc Nurs 2020 Feb 13;19(2):155-164. [doi: 10.1177/1474515119873648] [Medline: 31516036]
46. Rosenberg L, Kottorp A, Winblad B, Nygård L. Perceived difficulty in everyday technology use among older adults with or without cognitive deficits. Scand. J. of Occupational Therapy 2009:1-11. [doi: 10.3109/11038120802684299]
47. Stellefson M, Chaney B, Barry AE, Chavarria E, Tennant B, Walsh-Childers K, et al. Web 2.0 chronic disease self-management for older adults: a systematic review. J Med Internet Res 2013 Feb 14;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
48. Chung S, Nahm E. Testing reliability and validity of the eHealth Literacy Scale (eHEALS) for older adults recruited online. Comput Inform Nurs 2015 Apr;33(4):150-156 [FREE Full text] [doi: 10.1097/CIN.0000000000000146] [Medline: 25783223]
49. Li S, Yin Y, Cui G, Xu H. The Associations Among Health-Promoting Lifestyle, eHealth Literacy, and Cognitive Health in Older Chinese Adults: A Cross-Sectional Study. Int J Environ Res Public Health 2020 Mar 27;17(7):2263 [FREE Full text] [doi: 10.3390/ijerph17072263] [Medline: 32230935]
50. Zambianchi M, Carelli M. Attitudes Toward Technologies Questionnaire (ATTQ). Umeå, Sweden: Umeå University; 2013.
51. Davis F. Doctoral Dissertation: A technology acceptance model for empirically testing new end-user information systems: Theory and results. Massachusetts Institute of Technology. 1985. URL: https://dspace.mit.edu/handle/1721.1/15192 [accessed 2021-01-23]
52. Bear GG, Richards HC, Lancaster P. Attitudes toward Computers: Validation of a Computer Attitudes Scale. Journal of Educational Computing Research 1995 Jan 01;3(2):207-218. [doi: 10.2190/1dyt-1jej-t8j5-1yc7]
53. Schröder J, Sautier L, Kriston L, Berger T, Meyer B, Späth C, et al. Development of a questionnaire measuring Attitudes towards Psychological Online Interventions-the APOI. J Affect Disord 2015 Nov 15;187:136-141. [doi: 10.1016/j.jad.2015.08.044] [Medline: 26331687]
54. Loyd BH, Gressard C. Reliability and Factorial Validity of Computer Attitude Scales. Educational and Psychological Measurement 2016 Sep 07;44(2):501-505. [doi: 10.1177/0013164484442033]

55. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
56. Boot WR, Charness N, Czaja SJ, Sharit J, Rogers WA, Fisk AD, et al. Computer proficiency questionnaire: assessing low and high computer proficient seniors. Gerontologist 2015 Jun;55(3):404-411 [FREE Full text] [doi: 10.1093/geront/gnt117] [Medline: 24107443]
57. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
58. Edirippulige S. Changing Role of Nurses in the Digital Era: Nurses and Telehealth. In: Yogesan K, Bos L, Brett P, Gibbons MC, editors. Handbook of Digital Homecare. Cham, Switzerland: Springer Nature Switzerland AG; 2009:269-285.
59. Husebø AML, Storm M. Virtual visits in home health care for older adults. ScientificWorldJournal 2014;2014:689873-689811 [FREE Full text] [doi: 10.1155/2014/689873] [Medline: 25506616]
60. Savolainen L, Hanson E, Magnusson L, Gustavsson T. An Internet-based videoconferencing system for supporting frail elderly people and their carers. J Telemed Telecare 2008 Mar;14(2):79-82. [doi: 10.1258/jtt.2007.070601] [Medline: 18348753]
61. Munteanu C, Tennakoon C, Garner J, Goel A, Ho M, Shen C, et al. Improving older adults' online security: An exercise in participatory design. 2015 Presented at: Symposium on Usable Privacy and Security (SOUPS); July 22-24, 2015; Ottawa, Canada.
62. Collins SA, Currie LM, Bakken S, Vawdrey DK, Stone PW. Health literacy screening instruments for eHealth applications: a systematic review. J Biomed Inform 2012 Jun;45(3):598-607 [FREE Full text] [doi: 10.1016/j.jbi.2012.04.001] [Medline: 22521719]
63. Hu X, Bell RA, Kravitz RL, Orrange S. The prepared patient: information seeking of online support group members before their medical appointments. J Health Commun 2012;17(8):960-978. [doi: 10.1080/10810730.2011.650828] [Medline: 22574697]
64. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. J Clin Epidemiol 2011 Apr;64(4):366-379. [doi: 10.1016/j.jclinepi.2010.04.005] [Medline: 20638235]
65. Hsu MK, Wang SW, Chiu KK. Computer attitude, statistics anxiety and self-efficacy on statistical software adoption behavior: An empirical study of online MBA learners. Computers in Human Behavior 2009 Mar;25(2):412-420. [doi: 10.1016/j.chb.2008.10.003]
66. Weber B, Schneider B, Hornung S, Wetterling T, Fritze J. Computer attitude in psychiatric inpatients. Computers in Human Behavior 2008 Jul;24(4):1741-1752. [doi: 10.1016/j.chb.2007.07.006]
67. Ballance CT. Psychology of Computer Use: XXIV. Computer-Related Stress Among Technical College Students. Psychological Reports 1991 Oct;69(6):539. [doi: 10.2466/pr0.1991.69.2.539]
68. Kelley CL, Charness N. Issues in training older adults to use computers. Behaviour & Information Technology 1995 Mar;14(2):107-120. [doi: 10.1080/01449299508914630]
69. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
70. Sutherland LM, Kowalski KC, Ferguson LJ, Sabiston CM, Sedgwick WA, Crocker PR. Narratives of young women athletes' experiences of emotional pain and self-compassion. Qualitative Research in Sport, Exercise and Health 2014 Mar 14;6(4):499-516. [doi: 10.1080/2159676x.2014.888587]
71. Loos E. Senior citizens: Digital immigrants in their own country? Observatorio 2012;6(1):1-23 [FREE Full text]
72. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9]
73. Rodgers W, Herzog A. Collecting data about the oldest old: Problems and procedures. In: Manton KG, Willis DP, Suzman RM, editors. The oldest old. Oxford, England: Oxford University Press; 1995:135-156.

Abbreviations

APOI: Attitudes towards Psychological Online Interventions
ATC-IQ: Attitudes Toward Computer/Internet Questionnaire
CAS: Computer Attitude Scale
CCAT: Crowe Critical Appraisal Tool
DigComp: European Commission's Digital Competence framework
eHEALS: eHealth Literacy Scale
MDPQ: Mobile Device Proficiency Questionnaire
OECD: Organization for Economic Cooperation and Development
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

TAM: technology acceptance model
UTAUT: unified theory of acceptance and usage of technology

Edited by G Eysenbach; submitted 30.11.20; peer-reviewed by G Kolostoumpis; comments to author 24.12.20; revised version received 04.01.21; accepted 16.01.21; published 03.02.21.

Please cite as:
Oh SS, Kim KA, Kim M, Oh J, Chu SH, Choi J
Measurement of Digital Literacy Among Older Adults: Systematic Review
J Med Internet Res 2021;23(2):e26145
URL: https://www.jmir.org/2021/2/e26145
doi:10.2196/26145
PMID:33533727

Original Paper

Associations of Health Literacy, Social Media Use, and Self-Efficacy With Health Information–Seeking Intentions Among Social Media Users in China: Cross-sectional Survey

Zhaomeng Niu[1], PhD; Jessica Willoughby[2], PhD; Rongting Zhou[3], PhD

[1]Section of Behavioral Sciences, Division of Medical Oncology, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, United States

[2]The Edward R Murrow College of Communication, Washington State University, Pullman, WA, United States

[3]School of Humanities and Social Sciences, University of Science and Technology of China, Hefei, China

Corresponding Author:
Rongting Zhou, PhD
School of Humanities and Social Sciences
University of Science and Technology of China
Jinzhai Road 96
Hefei, 230026
China
Phone: 86 63600495
Email: rongting@ustc.edu.cn

Abstract

Background: Empirical research has demonstrated that people frequently use social media for gathering and sharing online health information. Health literacy, social media use, and self-efficacy are important factors that may influence people's health behaviors online.

Objective: We aimed to examine the associations between health literacy, health-related social media use, self-efficacy, and health behavioral intentions online.

Methods: We conducted a cross-sectional survey of adults 18 years and older (n=449) to examine predictors of health-related behavioral intentions online including health literacy, social media use, and self-efficacy in China using 2 moderated mediation models. Mediation and moderation analyses were conducted.

Results: Self-efficacy mediated the effects of health literacy ($B_{indirect}$=0.213, 95% CI 0.101 to 0.339) and social media use ($B_{indirect}$=0.023, 95% CI 0.008 to 0.045) on health behavioral intentions on social media. Age moderated the effects of health literacy on self-efficacy (*P*=.03), while previous experience moderated the effects of social media use on self-efficacy (*P*<.001).

Conclusions: Health literacy and health-related social media use influenced health behavioral intentions on social media via their prior effects on self-efficacy. The association between health literacy and self-efficacy was stronger among younger respondents, whereas the association between health-related social media use and self-efficacy was stronger among those who previously had positive experiences with health information on social media. Health practitioners should target self-efficacy among older populations and increase positive media experience related to health.

(J Med Internet Res 2021;23(2):e19134) doi:10.2196/19134

KEYWORDS

behavioral intention; health literacy; self-efficacy; social media

Introduction

The remarkably fast growth of the internet has made it a major source for information sharing and acquisition. In the late 1990s, the internet became a main source for health information [1]. Advantages of using the internet for health information include (1) it is the most convenient and comprehensive source; (2) the information seekers remain anonymous; and (3) it helps reduce inequalities and eliminate barriers (eg, distance) [2]. A study [3] found that, in China, 76.3% of computer-based users and 68.8% of mobile-based (eg, smartphone) users sought health information on the internet [3]. Furthermore, the advent of social media has enabled more possibilities such as connecting people

with similar health concerns or social support groups of patients [2,4].

With the rapid growth of social networking sites in China, the number of users and the variety of information on such sites have increased tremendously. As of 2018, there were 317 million active users of Weibo and 1 billion users of WeChat, which are the 2 main social media sites in China [5]. Weibo, a platform for microblogging, is often seen as the Chinese version of Twitter. WeChat is an instant messaging app that is similar to WhatsApp or Facebook Messenger, but it has more technological functions, such as free video or voice calls, group chat, public information sharing, mobile payments, and the ability to post pictures or videos [6].

Generally, extensive use of social media has been associated with informational or emotional need, professional development, social status, self-expression, and social interaction [7]. With an increasing awareness of health among the general public, a growing number of people in China are using social media for seeking and sharing health information [2,6]. Social media provide health information through multimedia affordances instead of solely text, which can increase the understanding of health information among populations with low health literacy [8]. Health literacy entails people's knowledge, motivation, and competency to access, understand, appraise, and apply health information in order to make judgments and decisions in everyday life concerning health care, disease prevention, and health promotion to maintain or improve their quality of life [9].

Additionally, a variety of health information and knowledge that used to be exclusive to health care providers are now available to health information seekers on social media [10]. Moreover, the user-generated nature of social media enables the sharing of health information and experiences, which provide views on health care from a patient's perspective and increase patient empowerment [11]. Approximately, 40% of the individuals who sought health information on social media also shared their personal health experiences [12]. Furthermore, previous studies [13,14] have shown that use of social media has positive effects on health behaviors.

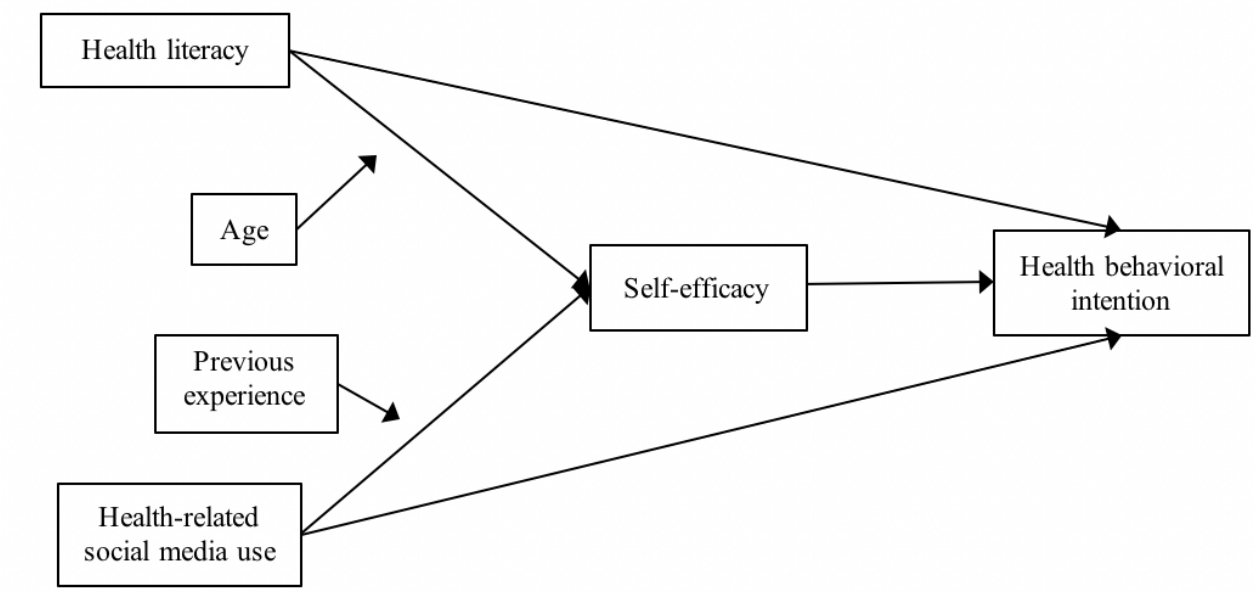
The health impact of social media in China has not been sufficiently studied and understood. For instance, research has focused on an examination of prominent health topics on social media [15] and how the general public views the impact of social media on health information acquisition [6]. Some studies have investigated the benefits and barriers [2], constructs of the theory of planned behaviors [16], cultural determinants, and doctor-patient communication [17] of health information intentions in China. Many studies in China have focused on health literacy among older adults or regionally [18-21]. However, no previous studies have comprehensively examined the potential relationships between health literacy, social media use, self-efficacy, and health behavioral intentions on social media in China based on the health literacy skills framework [22] and the social-cognitive perspective [23]. It is important to assess these constructs because studies have shown that health literacy levels in mainland China (the People's Republic of China excluding the special administrative regions of Hong Kong and Macau) remain low, which warrants further research regarding health literacy and related risk factors [24].

According to the health literacy skills framework, demographic factors, such as age, would moderate the development of health literacy while potential mediators, such as patient-provider communication [25] and knowledge [26], would mediate the effects of health literacy on health-related outcomes [22]. A number of studies [27-30] have found support for the mediating role of self-efficacy on health-related behavioral intentions [27,28]. Self-efficacy is the capacity to have positive effects on an individual's health [29]. Lee et al [30] found that self-efficacy mediated the effects of health literacy on health behaviors.

It has also been documented that social media use plays a positive role in health behaviors [14]. Previous studies [16,31] in China also found the association between social media use and health behavioral intentions. In addition, previous research has identified the association between media use and self-efficacy [32]. Social-cognitive theory posits that one's involvement with a subject grows over time through positive personal experiences, as one increases self-efficacy [33]. One study [34] found previous online health information seeking experience moderated seeking more of such information online. Therefore, health-related social media use and previous experience with using social media for health purposes could potentially influence self-efficacy and, consequently, impact health behavioral intentions.

The goal of this study was to examine the role of health literacy, social media use, and self-efficacy on health behavioral intentions on social media in China. Based on the literature, we proposed the following hypotheses (a conceptual model of the study is shown in Figure 1):

Figure 1. Conceptual framework.



- Hypothesis 1: Self-efficacy will mediate the effects of health literacy on health behavioral intentions on social media.
- Hypothesis 2: Self-efficacy will mediate the effects of health-related social media use on health behavioral intentions on social media.
- Hypothesis 3: Age will moderate the effects of health literacy on self-efficacy.
- Hypothesis 4: Previous experience with using social media for health will moderate the effects of health-related social media use on self-efficacy.

Methods

Recruitment

We disseminated an online survey on Sina Weibo (Sina Corporation) in China using a paid advertisement service. Participants were required to be social media users and be at least 18 years of age. The first page of the survey was an online consent form including study information. After reading the consent form, indicating that they agreed to participate and were 18 years or older, respondents were allowed to proceed. After completing the survey, respondents were entered in a raffle to win CNY ¥300 (approximately US $42.86). The study was reviewed by the university institutional review board and was approved as an exempt study.

Measures

Health-related social media use was measured by one question asking how frequently they have used social media for health information before on a scale ranging from 1 (never) to 7 (multiple times a day) (mean 4.01, SD 1.86). An English instructor at a Chinese university translated the questionnaire and used back-translation to ensure consistency in wording between English and Chinese versions of the survey.

Health literacy was measured by a scale adapted from Chinn and McCarthy [35] for health literacy measurement. This scale included seven 3-point items (rarely, sometimes, often), such as “When you talk to a doctor or nurse, do you give them all the information they need to help you?” and “Are you someone who likes to find out lots of different information about your health?” (mean 2.14, SD 0.40; Cronbach α=.71).

Self-efficacy was captured by the self-efficacy scale from Lee et al [29]. Five 7-point Likert-type items measured the degree of agreement with statements regarding self-efficacy in managing one’s health, such as “I have been able to meet the goals I set for myself to improve my health” and “I am confident I can have a positive effect on my health” (mean 5.04, SD 1.04; Cronbach α=.85).

Previous experience was measured by asking the respondents whether they found the health information on social media useful on three 7-point Likert scale items such as “In the past three months, health advice offered on social media sites has been useful to me” (mean 4.01, SD 1.58; Cronbach α=.91).

Behavioral intention was measured from an adapted multidimensional scale [36]. Three 7-point Likert-type items asked the extent to which respondents agree with the statements about their behavioral intention regarding health information on social media including “I will act upon the advice that is offered in the message in the near future,” “I will forward the message to my online acquaintances,” and “I will recommend the advice I read in the message to another person” (mean 3.84, SD 1.42; Cronbach α=.90).

We also measured demographic variables including age, sex, education, and family yearly income. Sex, education, and family yearly income were measured with 3 categorical questions while age was measured by asking participants to indicate their age in numbers (Multimedia Appendix 1).

Statistical Analysis

To test the hypothesized associations, we used estimated direct and indirect effects in mediation and moderation models using SPSS statistical software (version 25.0, IBM Corp; PROCESS macro [37]). Mediation models (PROCESS model 4) were used

for hypothesis 1 and hypothesis 2, whereas PROCESS model 7 was used to test the moderated mediation effect. Age, sex, education, and family yearly income were controlled as covariates for hypotheses 1, 2, and 4, whereas sex, education, and family yearly income were controlled as covariates for hypothesis 3.

Results

User Statistics

A total of 608 respondents began the questionnaire; however, 127 were removed due to declining to participate or incomplete participation (defined as more than 50% of the survey not completed), and 32 were excluded due to missing data. We had a final sample size of 449 (women: n=345; men: n=104; age: mean 25.23 years, SD 5.23, range 18-66). There were 242 participants aged from 23 to 30 years old (242/449, 53.9%). The education level of the sample was high, with 52.6% of the respondents (236/449) reporting having a bachelor's degree, and 53.7% of the respondents (241/449) reported their family annual income was between ¥50,000 (approximately US $7150) to ¥200,000 (approximately US $28,600) (Table 1).

Table 1. Descriptive statistics.

Variable	Value
Behavioral intentions, mean (SD)	3.84 (1.42)
Health literacy, mean (SD)	2.16 (.42)
Self-efficacy, mean (SD)	5.04 (1.04)
Social media use, mean (SD)	3.80 (1.48)
Previous experience, mean (SD)	3.99 (1.60)
Gender, n (%)	
Female	345 (76.8)
Male	104 (23.2)
Age (years), mean (SD)	25.23 (5.23)
Education, n (%)	
High school degree or lower	16 (3.6)
College degree or some college	236 (52.6)
Graduate degree or higher	197 (43.8)
Family yearly income (CNY[a]), n (%)	
0-50,000	124 (27.6)
50,001-100,000	112 (25.0)
100,001-200,000	129 (28.7)
>200,001	84 (18.7)

[a]An approximate exchange rate of 1 CNY to US $0.143 is applicable.

Evaluation Outcomes

According to the results of model 4, self-efficacy mediated the effects of health literacy on health behavioral intentions on social media ($B_{indirect}$=0.213, SE 0.060, 95% CI 0.101 to 0.339). Thus, hypothesis 1 was supported. The participants who had higher health literacy also had higher self-efficacy and then would be more likely to intend to perform health behaviors based on information acquired on social media (eg, use the health advice they found). Self-efficacy mediated the effects of health-related social media use on health behavioral intentions on social media ($B_{indirect}$=0.023, SE 0.009, 95% CI 0.008 to 0.045), indicating hypothesis 2 was supported—the more the participants used social media for health, the higher their self-efficacy, and they would be more likely to have greater intentions to perform health behaviors on social media.

Moderated Mediation

The results of moderated mediation models are shown in Table 2 and Table 3. According to the results of model 7, age moderated the effects of health literacy on self-efficacy (B=–0.041, SE 0.019, P=.03). The interaction had a negative effect on self-efficacy. Thus, hypothesis 3 was supported. The moderated mediation model accounted for 10.7% variance in health behavioral intention. Health literacy had a direct effect (B=0.345, SE 0.141, P=.02) as well as an indirect effect (Table 4) on health behavioral intention.

Previous experience with using social media for health moderated the effects of health-related social media use on self-efficacy (B=0.058, SE 0.015, P<.001). The interaction had a positive effect on self-efficacy. Therefore, hypothesis 4 was supported. This moderated mediation model accounted for 31.3% variance in health behavioral intention. Health-related

social media use had a direct effect (B=0.315, SE 0.0309, P<.001) as well as an indirect effect (Table 5) on health behavioral intention.

Table 2. Regression results for effects of self-efficacy, age, and health literacy on health behavioral intention (R^2=0.107, P<.001).

Variable	B^a (SE)	P value	95% CI
Sex	0.156 (0.152)	.30	–0.142 to 0.454
Education	0.214 (0.086)	.01	0.046 to 0.382
Family yearly income	0.141 (0.043)	.001	0.058 to 0.225
Health literacy	0.345 (0.141)	.01	0.068 to 0.621
Health literacy × age on self-efficacy	–0.041 (0.019)	.03	–0.077 to–0.004
Self-efficacy	0.318 (0.069)	<.001	0.183 to 0.453
Direct effect of health literacy on health behavioral intention	0.345 (0.141)	.01	0.068 to 0.621
Index of moderated mediation: age	–0.013 (0.006)	N/A[b]	–0.025 to –0.001

[a]Unstandardized final model coefficients.

[b]N/A: not applicable.

Table 3. Regression results for effects of self-efficacy, previous experience, and health-related social media use on health behavioral intention (R^2=0.313, P<.001)

Variable	B^a (SE)	P value	95% CI
Age	0.052 (0.011)	<.001	0.029 to 0.074
Sex	0.138 (0.136)	.31	–0.128 to 0.405
Education	0.259 (0.078)	.001	0.106 to 0.412
Family yearly income	0.145 (0.037)	<.001	0.071 to 0.218
Health-related social media use	0.315 (0.031)	<.001	0.254 to 0.376
Health-related social media use × previous experience on self-efficacy	0.058 (0.015)	<.001	0.029 to 0.087
Self-efficacy	0.250 (0.057)	<.001	0.137 to 0.363
Direct effect of health-related social media use on health behavioral intention	0.315 (0.031)	<.001	0.254 to 0.376
Index of moderated mediation: previous experience	0.015 (0.005)	N/A[b]	0.006 to 0.026

[a]Unstandardized final model coefficients.

[b]N/A: not applicable.

Table 4. Conditional indirect effects of health literacy on health behavioral intentions by age.

Age (years)	Effect	SE	95% CI
20	0.294	0.079	0.145 to 0.452
25	0.226	0.062	0.109 to 0.351
30	0.158	0.059	0.061 to 0.290

Table 5. Conditional indirect effects of health-related social media use on health behavioral intentions by previous experience.

Previous experience (score)	Effect	SE	95% CI
mean – 1 SD (=2.414)	–0.017	0.012	–0.044 to 0.002
mean (=4.005)	0.006	0.007	0.009 to 0.020
mean + 1 SD (=5.597)	0.029	0.010	0.011 to 0.050

Moderation Effects

Conditional indirect effects of health literacy on health behavioral intentions by age are shown in Table 4, and conditional indirect effects of health-related social media use on health behavioral intentions by previous experience are shown in Table 5. The positive indirect relationship between health literacy on health behavioral intentions was stronger

among the younger segment of our sample (point estimate 0.294, SE 0.079, 95% CI 0.145 to 0.452). Additionally, the positive indirect effect of health-related social media use on health behavioral intentions was stronger among those participants who previously had positive experience with health information on social media, who were at one standard deviation above the mean (point estimate 0.029, SE 0.010, 95% CI 0.011 to 0.050). The moderated mediation results also revealed that self-efficacy remained a significant mediator no matter whether the participants were at 20, 25, or 30 years old. However, self-efficacy was only a significant mediator when the participants had the mean score of previous experience with health information on social media or one standard deviation above the mean score.

Discussion

Principal Findings

A substantial number of studies have examined health information on social media in China [6,16-18,28,38]; however, no previous studies have examined health behavioral intentions on a Chinese social media site from both the health literacy skills framework [22] or social-cognitive perspectives [23]. And there has been no study to comprehensively investigate how health information with different features on social media influenced the trust in such health information. Some studies [39,40] have examined the role of past experience and social media use individually, and some studies [18,20,21] have tested the associations between self-efficacy and other health outcomes; however, there are few empirical studies examining the relationship among health literacy, past experience, health-related social media use, and health behavioral intention on social media.

Our findings indicate that health literacy and health-related social media use influenced health behavioral intentions both directly and through their prior effects on self-efficacy in managing one's health. Individuals with higher levels of health literacy had greater self-efficacy in managing their health and then, consequently, had greater health behavioral intentions on social media such as using the health information they found or sharing health information with others from the internet. Part of this finding is also consistent with those from previous studies [9,41,42] suggesting that health literacy is positively associated with information sharing. We also found that individuals who used social media for health purposes more frequently were more likely to report higher self-efficacy in managing their health and greater health behavioral intentions on social media. Social media usage for information could improve people's psychological state and increase confidence and motivations to cope with uncertainties [43]. Therefore, people who use social media for health more frequently would be able to learn new information, cases, health experiences of others, and avoid potential risks, which could lead to a higher confidence in managing one's own health. Our finding regarding the positive association between self-efficacy and health behavioral intentions is consistent with the social-cognitive theory perspective and previous empirical studies [44,45] examining effects of self-efficacy on different health behaviors and behavioral intentions.

Another important finding from our study pertains to the moderated mediation effects. Higher health literacy was associated with greater self-efficacy in health, which in turn was related to higher health behavioral intentions on social media. This relationship between health literacy and self-efficacy was moderated by age, suggesting that health literacy increased self-efficacy among younger social media users and eventually promoted their health behavioral intentions on social media. Younger social media users who had greater health literacy tended to have higher confidence in managing their own health and consequently had greater intentions to perform health behaviors. Among the older segment of social media users, no matter their health literacy level, their confidence in managing, improving, and generating positive effects on their health was lower than those of the younger groups. A number of studies [46-48] have found the negative association between age and self-efficacy. In one study [49], older adults who had lower incomes and lower education had relatively low self-efficacy, which was similar to our findings regarding the results of socioeconomic status and self-efficacy.

The positive relationship between health-related social media use and self-efficacy was stronger among those who had previously benefited from using social media for health. When people seek health information online, they usually not only experience increases in health knowledge but also find social support and help from people in similar situations [50]. Therefore, individuals with prior positive experiences using social media for health would have greater efficacy in exerting positive effects on one's health. Previous experience with applying health advice found on social media in real life that resulted in good health results would improve their confidence in continuing to seek and use health information online.

Implications

With the rapid growth of social media use, this study has important implications for health practitioners. A framework for health behavioral intentions was constructed based on the components of the health literacy skills framework [22] and the social-cognitive perspective [23].

Health literacy influenced self-efficacy and health behavioral intentions, which highlights the importance of health literacy level in China. The concept of health literacy is not popular in China and the quality of medical services provided in China varies significantly based on areas. This makes it important to improve the health literacy level in China so that people can have the ability to take effective and accurate actions related to health. Given the moderating role of age, participants between the ages of 25 and 30 years in our sample require more customized interventions, such as including carefully evaluated digital elements [51,52], to improve their self-efficacy in managing their health in health interventions.

Greater health-related use of social media was associated with higher self-efficacy and health behavioral intentions, indicating the importance of social media in understanding health behavioral intentions. Since the association between

health-related social media use and self-efficacy was increased by positive experience with social media for health, health practitioners and scholars should aim to improve users' experiences with social media regarding health information.

The findings are important for health scholars interested in understanding the factors that influence the intent to use health information on social media sites. This study also provides insights for health message designers who want to build effective health campaigns and distribute accurate and credible health information on social media platforms. Future studies should explicitly investigate how to improve health literacy levels and users' experience with social media, such as by developing health literacy education programs [53].

Limitations

Limitations in this study should be considered. First, we used a convenience sample on social media. The service used to advertise the survey link claimed to spread the survey post randomly, however, those who were interested in this study might share some similar traits (such as being in a younger population group or interested in this topic). Therefore, the sample was not truly representative of social media users in general, which might limit generalizability to other populations. Our sample was biased toward younger populations. Future studies could use different means to distribute the survey in order to reach a more diverse audience.

Second, while we asked people if they would act on the health information, we cannot verify or assess the potential accuracy of information that would be obtained. Future work should also consider credibility of the sources and information presented as part of the findings in terms of whether acting on information would be beneficial for health [54], especially in the online environment in which health misinformation may be rampant.

Finally, this study focused on investigating whether or not social media use could predict health-related behavioral intentions. Although previous research suggested that social media use could be influenced by cognition and behaviors [55], we did not test the reinforcing spiral framework of social media use in this study. This framework indicates that media use can influence attitudinal or behavioral outcomes, which can in turn affect habits of using media. According to this framework, media use can be an outcome of psychological processing and behaviors and also can influence psychological and behavioral results. Future studies should examine the reinforcing role of social media use in predicting health-related behavioral intentions on social media.

Conclusions

Health literacy and health-related social media use in China participants influenced health behavioral intentions on social media via their prior effects on self-efficacy in health. The association between health literacy and self-efficacy was stronger among younger respondents, whereas the association between health-related social media use and self-efficacy was stronger among those who previously had positive prior experience with health information on social media. Our results provide insights for health practitioners and researchers and increase understanding of the mechanisms behind using social media for health.

Acknowledgments

The authors would like to thank Dr Bo Li and Dr Ze Li for their insightful feedback on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Variables.
[DOCX File , 29 KB - jmir_v23i2e19134_app1.docx]

References

1. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004 Nov;59(9):1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
2. Li Y, Wang X, Lin X, Hajli M. Seeking and sharing health information on social media: a net valence model and cross-cultural comparison. Technological Forecasting and Social Change 2018 Jan;126:28-40. [doi: 10.1016/j.techfore.2016.07.021]
3. Di Z, Junsheng G, Ruosi S. Clustering analysis of health information acquisition channels: active acquisition and passive receiving. J Int Commun 2015;5:81-93.
4. Oh S, Syn SY. Motivations for sharing information and social support in social media: a comparative analysis of Facebook, Twitter, Delicious, YouTube, and Flickr. J Assn Inf Sci Tec 2015 Apr 21;66(10):2045-2060. [doi: 10.1002/asi.23320]
5. Zhu Y. Social media engagement and Chinese international student recruitment: understanding how UK HEIs use Weibo and WeChat. Journal of Marketing for Higher Education 2019 Jun 23;29(2):173-190. [doi: 10.1080/08841241.2019.1633003]
6. Zhang X, Wen D, Liang J, Lei J. How the public uses social media wechat to obtain health information in china: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]

7. Zhang L, Pentina I. Motivations and usage patterns of Weibo. Cyberpsychol Behav Soc Netw 2012 Jun;15(6):312-317. [doi: 10.1089/cyber.2011.0615] [Medline: 22703037]
8. Roberts M, Callahan L, O'Leary C. Social media: a path to health literacy. ISU 2017 Jun 26;37(2):177-187. [doi: 10.3233/isu-170836]
9. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12(1):80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
10. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
11. Verhoef LM, Van de Belt TH, Engelen LJ, Schoonhoven L, Kool RB. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014 Feb 20;16(2):e56 [FREE Full text] [doi: 10.2196/jmir.3024] [Medline: 24566844]
12. Fox S, Duggan M. Health online 2013. Pew Research Center. Washington DC URL: https://www.ordinedeimedici.com/documenti/Docs7-cybercondria-PIP-HealthOnline.pdf [accessed 2018-06-06]
13. Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. Health Info Libr J 2007 Mar;24(1):2-23 [FREE Full text] [doi: 10.1111/j.1471-1842.2007.00701.x] [Medline: 17331140]
14. Laranjo L, Arguel A, Neves A, Gallagher A, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256 [FREE Full text] [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
15. Wang S, Paul M, Dredze M. Exploring health topics in Chinese social media: an analysis of Sina Weibo. 2014 Presented at: Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence; Jun 18; Québec, Canada.
16. Niu Z, Willoughby JF. Examining cultural identity and media use as predictors of intentions to seek mental health information among Chinese. Asian Journal of Communication 2018 Feb 17;28(4):360-376. [doi: 10.1080/01292986.2018.1441319]
17. Wu T, Deng Z, Zhang D, Buchanan PR, Zha D, Wang R. Seeking and using intention of health information from doctors in social media: the effect of doctor-consumer interaction. Int J Med Inform 2018 Jul;115:106-113. [doi: 10.1016/j.ijmedinf.2018.04.009] [Medline: 29779712]
18. Liu Y, Liu L, Li Y, Chen Y. Relationship between health literacy, health-related behaviors and health status: a survey of elderly Chinese. Int J Environ Res Public Health 2015 Aug 18;12(8):9714-9725 [FREE Full text] [doi: 10.3390/ijerph120809714] [Medline: 26295246]
19. Wu Y, Wang L, Cai Z, Bao L, Ai P, Ai Z. Prevalence and risk factors of low health literacy: a community-based study in Shanghai, China. Int J Environ Res Public Health 2017 Jun 12;14(6):628 [FREE Full text] [doi: 10.3390/ijerph14060628] [Medline: 28604645]
20. Liu Y, Xue L, Xue H, Hou P. Health literacy, self-care agency, health status and social support among elderly Chinese nursing home residents. Health Education Journal 2017 Nov 17;77(3):303-311. [doi: 10.1177/0017896917739777]
21. Yang Y, Zhang B, Meng H, Liu D, Sun M. Mediating effect of social support on the associations between health literacy, productive aging, and self-rated health among elderly Chinese adults in a newly urbanized community. Medicine 2019;98(16):e15162. [doi: 10.1097/md.0000000000015162]
22. Squiers L, Peinado S, Berkman N, Boudewyns V, McCormack L. The health literacy skills framework. J Health Commun 2012 Oct;17 Suppl 3(sup3):30-54. [doi: 10.1080/10810730.2012.713442] [Medline: 23030560]
23. Bandura A. Social cognitive theory of mass communication. In: Media Effects. New York: Routledge; 2009:110-140.
24. Wang C, Li H, Li L, Xu D, Kane RL, Meng Q. Health literacy and ethnic disparities in health-related quality of life among rural women: results from a Chinese poor minority area. Health Qual Life Outcomes 2013 Sep 11;11(1):153 [FREE Full text] [doi: 10.1186/1477-7525-11-153] [Medline: 24020618]
25. Schillinger D. Improving the quality of chronic disease management for populations with low functional health literacy: a call to action. Disease Management 2001 Sep;4(3):103-109. [doi: 10.1089/10935070152596025]
26. Lee SD, Arozullah AM, Cho YI. Health literacy, social support, and health: a research agenda. Social Science & Medicine 2004 Apr;58(7):1309-1321. [doi: 10.1016/s0277-9536(03)00329-0]
27. Cao W, Zhang X, Xu K, Wang Y. Modeling online health information-seeking behavior in China: the roles of source characteristics, reward assessment, and internet self-efficacy. Health Commun 2016 Sep 09;31(9):1105-1114. [doi: 10.1080/10410236.2015.1045236] [Medline: 26861963]
28. Deng Z, Liu S. Understanding consumer health information-seeking behavior from the perspective of the risk perception attitude framework and social support in mobile social media websites. Int J Med Inform 2017 Sep;105:98-109. [doi: 10.1016/j.ijmedinf.2017.05.014] [Medline: 28750916]
29. Sun Young Lee, Hwang H, Hawkins R, Pingree S. Interplay of negative emotion and health self-efficacy on the use of health information and its outcomes. Communication Research 2008 Apr 14;35(3):358-381. [doi: 10.1177/0093650208315962]
30. Lee JY, Divaris K, Baker AD, Rozier RG, Vann WF. The relationship of oral health literacy and self-efficacy with oral health status and dental neglect. Am J Public Health 2012 May;102(5):923-929. [doi: 10.2105/ajph.2011.300291]

31. Niu Z, Willoughby JF, Mei J, Li S, Hu P. A cross-cultural comparison of an extended planned risk information seeking model on mental health among college students: cross-sectional study. J Med Internet Res 2020 May 11;22(5):e15817 [FREE Full text] [doi: 10.2196/15817] [Medline: 32441654]
32. Kushin M, Yamamoto M. Did social media really matter? college students' use of online media and political decision making in the 2008 election. Mass Communication and Society 2010 Oct 29;13(5):608-630. [doi: 10.1080/15205436.2010.516863]
33. Bandura A. Self-efficacy and health behaviour. Cambridge Handbook of Psychology, Health and Medicine 1997:160-162.
34. Jiang S, Street RL. Pathway linking internet health information seeking to better health: a moderated mediation studyPathway Linking Internet Health Information Seeking to Better Health: A Moderated Mediation Study. Health Commun 2017 Aug 27;32(8):1024-1031. [doi: 10.1080/10410236.2016.1196514] [Medline: 27464036]
35. Chinn D, McCarthy C. All Aspects of Health Literacy Scale (AAHLS): developing a tool to measure functional, communicative and critical health literacy in primary healthcare settings. Patient Educ Couns 2013 Feb;90(2):247-253. [doi: 10.1016/j.pec.2012.10.019] [Medline: 23206659]
36. Hu Y, Shyam Sundar S. Effects of online health sources on credibility and behavioral intentions. Commun Res 2009 Nov 25;37(1):105-132. [doi: 10.1177/0093650209351512]
37. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York: Guilford Publications; 2017.
38. Lee JY, Sundar SS. To tweet or to retweet? that is the question for health professionals on twitter. Health Commun 2013 Aug 08;28(5):509-524. [doi: 10.1080/10410236.2012.700391] [Medline: 22873787]
39. Ho SS, Detenber BH, Rosenthal S, Lee EWJ. Seeking information about climate change. Sci Commun 2014 Feb 07;36(3):270-295. [doi: 10.1177/1075547013520238]
40. Hovick SR, Kahlor L, Liang M. Personal cancer knowledge and information seeking through PRISM: the planned risk information seeking model. J Health Commun 2014 Apr 16;19(4):511-527. [doi: 10.1080/10810730.2013.821556] [Medline: 24433251]
41. Crook B, Stephens KK, Pastorek AE, Mackert M, Donovan EE. Sharing health information and influencing behavioral intentions: the role of health literacy, information overload, and the internet in the diffusion of healthy heart information. Health Commun 2016 Feb 10;31(1):60-71. [doi: 10.1080/10410236.2014.936336] [Medline: 25668744]
42. Ledford CJW, Cafferty LA, Russell TC. The influence of health literacy and patient activation on patient information seeking and sharing. J Health Commun 2015;20 Suppl 2:77-82. [doi: 10.1080/10810730.2015.1066466] [Medline: 26513034]
43. Hu S, Gu J, Zhang S. Social media usage, self-efficacy and cultural intelligence: a longitudinal empirical research in China. 2017 Presented at: Wuhan International Conference on e-Business. Association For Information Systems; May 26; Wuhan, China.
44. Schwarzer R, Schüz B, Ziegelmann JP, Lippke S, Luszczynska A, Scholz U. Adoption and maintenance of four health behaviors: theory-guided longitudinal studies on dental flossing, seat belt use, dietary behavior, and physical activity. Ann Behav Med 2007 Jun;33(2):156-166. [doi: 10.1007/bf02879897]
45. Gao Z, Xiang P, Lee AM, Harrison L. Self-efficacy and outcome expectancy in beginning weight training class: their relations to students' behavioral intention and actual behavior. Res Quart Exerc Sport 2008 Mar 01;79(1):92-100. [doi: 10.5641/193250308x13086753543059]
46. Clark DO. Age, socioeconomic status, and exercise self-efficacy. Gerontologist 1996 Apr 01;36(2):157-164. [doi: 10.1093/geront/36.2.157] [Medline: 8920084]
47. Baker DW, Gazmararian JA, Sudano J, Patterson M. The association between age and health literacy among elderly persons. J Gerontol B Psychol Sci Soc Sci 2000 Nov 01;55(6):S368-S374. [doi: 10.1093/geronb/55.6.s368] [Medline: 11078114]
48. Kobayashi LC, Wardle J, Wolf MS, von Wagner C. Aging and functional health literacy: a systematic review and meta-analysis. J Gerontol B Psychol Sci Soc Sci 2016 May;71(3):445-457 [FREE Full text] [doi: 10.1093/geronb/gbu161] [Medline: 25504637]
49. Gecas V. The social psychology of self-efficacy. Annu Rev Sociol 1989 Aug;15(1):291-316. [doi: 10.1146/annurev.so.15.080189.001451]
50. Niu Z, Jeong DC, Willoughby JF. Friends over doctors? the influences of source and perceived customization on college drinking. Health Commun 2020 Jan 10:1-11. [doi: 10.1080/10410236.2020.1712034] [Medline: 31920113]
51. Rosenberg D, Depp CA, Vahia IV, Reichstadt J, Palmer BW, Kerr J, et al. Exergames for subsyndromal depression in older adults: a pilot study of a novel intervention. The American Journal of Geriatric Psychiatry 2010 Mar;18(3):221-226. [doi: 10.1097/jgp.0b013e3181c534b5]
52. Seifert A, Reinwand DA, Schlomann A. Designing and using digital mental health interventions for older adults: being aware of digital inequality. Front Psychiatry 2019;10:568. [doi: 10.3389/fpsyt.2019.00568] [Medline: 31447716]
53. Kim H, Xie B. Health literacy in the eHealth era: a systematic review of the literature. Patient Educ Couns 2017 Jun;100(6):1073-1082. [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]
54. Braithwaite DO, Waldron VR, Finn J. Communication of social support in computer-mediated groups for people with disabilities. Health Commun 1999 Apr;11(2):123-151. [doi: 10.1207/s15327027hc1102_2] [Medline: 16370973]

55. Slater MD. Reinforcing spirals: the mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. Commun Theory 2007 Aug;17(3):281-303. [doi: 10.1111/j.1468-2885.2007.00296.x]

Edited by G Eysenbach; submitted 10.06.20; peer-reviewed by N Daniëls, JR Bautista, C Mather; comments to author 01.07.20; revised version received 04.09.20; accepted 26.10.20; published 25.02.21.

Please cite as:
Niu Z, Willoughby J, Zhou R
Associations of Health Literacy, Social Media Use, and Self-Efficacy With Health Information–Seeking Intentions Among Social Media Users in China: Cross-sectional Survey
J Med Internet Res 2021;23(2):e19134
URL: https://www.jmir.org/2021/2/e19134
doi:10.2196/19134
PMID:33629955

Original Paper

A Bespoke Electronic Health Record for Epilepsy Care (EpiToMe): Development and Qualitative Evaluation

Shiqiang Tao[1,2], PhD; Samden Lhatoo[1,2], MD, FCAP; Johnson Hampson[1,2], MS; Licong Cui[2,3], PhD; Guo-Qiang Zhang[1,2,3], PhD

[1]Department of Neurology, The University of Texas Health Science Center at Houston, Houston, TX, United States

[2]Texas Institute for Restorative Neurotechnologies, The University of Texas Health Science Center at Houston, Houston, TX, United States

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

Corresponding Author:
Guo-Qiang Zhang, PhD
Department of Neurology
The University of Texas Health Science Center at Houston
1133 John Freeman Blvd
JJL 430
Houston, TX, 77030
United States
Phone: 1 7135007117
Email: guo-qiang.zhang@uth.tmc.edu

Abstract

Background: While electronic health records (EHR) bring various benefits to health care, EHR systems are often criticized as cumbersome to use, failing to fulfill the promise of improved health care delivery with little more than a means of meeting regulatory and billing requirements. EHR has also been recognized as one of the contributing factors for physician burnout.

Objective: Specialty-specific EHR systems have been suggested as an alternative approach that can potentially address challenges associated with general-purpose EHRs. We introduce the Epilepsy Tracking and optimized Management engine (EpiToMe), an exemplar bespoke EHR system for epilepsy care. EpiToMe uses an agile, physician-centered development strategy to optimize clinical workflow and patient care documentation. We present the design and implementation of EpiToMe and report the initial feedback on its utility for physician burnout.

Methods: Using collaborative, asynchronous data capturing interfaces anchored to a domain ontology, EpiToMe distributes reporting and documentation workload among technicians, clinical fellows, and attending physicians. Results of documentation are transmitted to the parent EHR to meet patient care requirements with a push of a button. An HL7 (version 2.3) messaging engine exchanges information between EpiToMe and the parent EHR to optimize clinical workflow tasks without redundant data entry. EpiToMe also provides live, interactive patient tracking interfaces to ease the burden of care management.

Results: Since February 2019, 15,417 electroencephalogram reports, 2635 Epilepsy Monitoring Unit daily reports, and 1369 Epilepsy Monitoring Unit phase reports have been completed in EpiToMe for 6593 unique patients. A 10-question survey was completed by 11 (among 16 invited) senior clinical attending physicians. Consensus was found that EpiToMe eased the burden of care documentation for patient management, a contributing factor to physician burnout.

Conclusions: EpiToMe offers an exemplar bespoke EHR system developed using a physician-centered design and latest advancements in information technology. The bespoke approach has the potential to ease the burden of care management in epilepsy. This approach is applicable to other clinical specialties.

(J Med Internet Res 2021;23(2):e22939) doi:10.2196/22939

KEYWORDS

specialty-specific EHR; physician-centered design; clinical workflow; patient care management; clinical care documentation; physician burnout; interoperability

Introduction

Electronic Health Records

Electronic health records (EHR) have been broadly adopted in the United States in the last 2 decades to improve the quality of health care, increase patient satisfaction, and save health care costs [1-3], as mandated by the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 [4-7]. Compared to paper-based medical records, EHR has advantages that include easier access, higher working efficiency, increased patient satisfaction, reduced financial cost, better data exchange and interoperability, and opportunities for secondary use of clinical data for research [8-11].

While EHR brings such benefits to health care, EHR systems are often criticized as cumbersome to use, failing to fulfill the promise of improved health care delivery with little more than a means of meeting regulatory and billing requirements [12]. A recent study inspected the time allocation pattern among over 31 million transactions for 471 physicians from 2011 to 2014 and found that physicians spent progressively more time on “desktop medicine” and less on face-to-face patient care [13]. Another study inspected EHR event logs and showed that primary care physicians spend more than half of their workday interacting with the EHR during and after clinic hours [14].

EHR systems have been recognized as one of the contributing factors for physician burnout [15,16], an increasing health care crisis in the United States [17-20]. Burnout is on the rise and affects all specialties [21]. Studies show that burnt-out doctors are more likely to make medical errors [22], work less efficiently [23], and have higher referral rates [24]. A recent survey of nearly 6880 physicians reported that 1 in 50 planned to leave medicine altogether in the next 2 years, while 1 in 5 planned to reduce clinical hours over the next year [25]. Another study [26] reported that 26% of 1792 physician respondents reported burnout, and 70% of 1631 users reported EHR-related stress. The study also reported that high rates of fatigue among intensive care unit physicians were associated with low EHR efficiency [27].

Specialty-Specific EHRs

One recent study pointed out that different specialties had different unique requirements, and this difference should be reflected in EHR design and implementation [28]. Specialty-specific or bespoke EHR is a promising approach to overcoming the limitations of general-purpose EHR and mitigating physician burnout. A bespoke EHR is an EHR custom designed to meet the unique needs of providers in a specific specialty or care setting. Bespoke EHR can prevent clinicians from spending a significant portion of their workday sifting through large amounts of clinical data for the specific data elements they need. In another recent study [29], it was reported that a clinic-focused Sprint process can optimize EHR efficiency and have positive effects on physician burnout. Specialty-specific EHR improvement is one major intervention during the Sprint process. In general, specialty-specific EHR can better achieve the level of optimization and workflow management expected by physicians [30], although approaches based on EHR customization have limitations in what is achievable compared to a bespoke design built from the ground up. Standalone, specialty-specific EHRs have been around for a number of years in such areas such as emergency medicine, ophthalmology, and dermatology. However, broader adoption of such a specialty-specific approach faces challenges in interoperability between different EHR systems, capturing standardized structured data for documenting care, and supporting the data-readiness needs to drive a learning health system.

EpiToMe: A Bespoke EHR

We developed EpiToMe (an Epilepsy Tracking and optimization Management engine), a bespoke EHR system customized for epilepsy care created de novo. EpiToMe has evolved from and integrates clinical applications we have developed over the last decade [31-34]. EpiToMe provides patient data capture functions for electroencephalogram (EEG) reporting, daily reporting, and phase reporting for Epilepsy Monitoring Units (EMUs). It uses domain-specific epilepsy and seizure ontology (EpSO) [35] to (1) support structured entry of multimodal epilepsy data, (2) proactively ensure the quality of data through the use of ontology terms in faceted systems, (3) organize and index patient information for subsequent analytical queries and secondary use, and (4) seamlessly make just-in-time and right-in-context communications with the parent EHR. EpiToMe was developed following web interface–driven development [33], an agile software development methodology, in close collaboration with physicians. EpiToMe has a built-in physician dashboard optimized for physician needs to perform tasks without switching systems or changing navigation interfaces. EpiToMe's data entry pipeline allows other clinicians in the team such as EEG technicians and clinical fellows to take responsibility for appropriate patient data documentation work. EpiToMe also provides a tracker to provide an overview of patient status in the clinical care workflow.

Methods

Physician-Centered Design and Interface-Driven Development

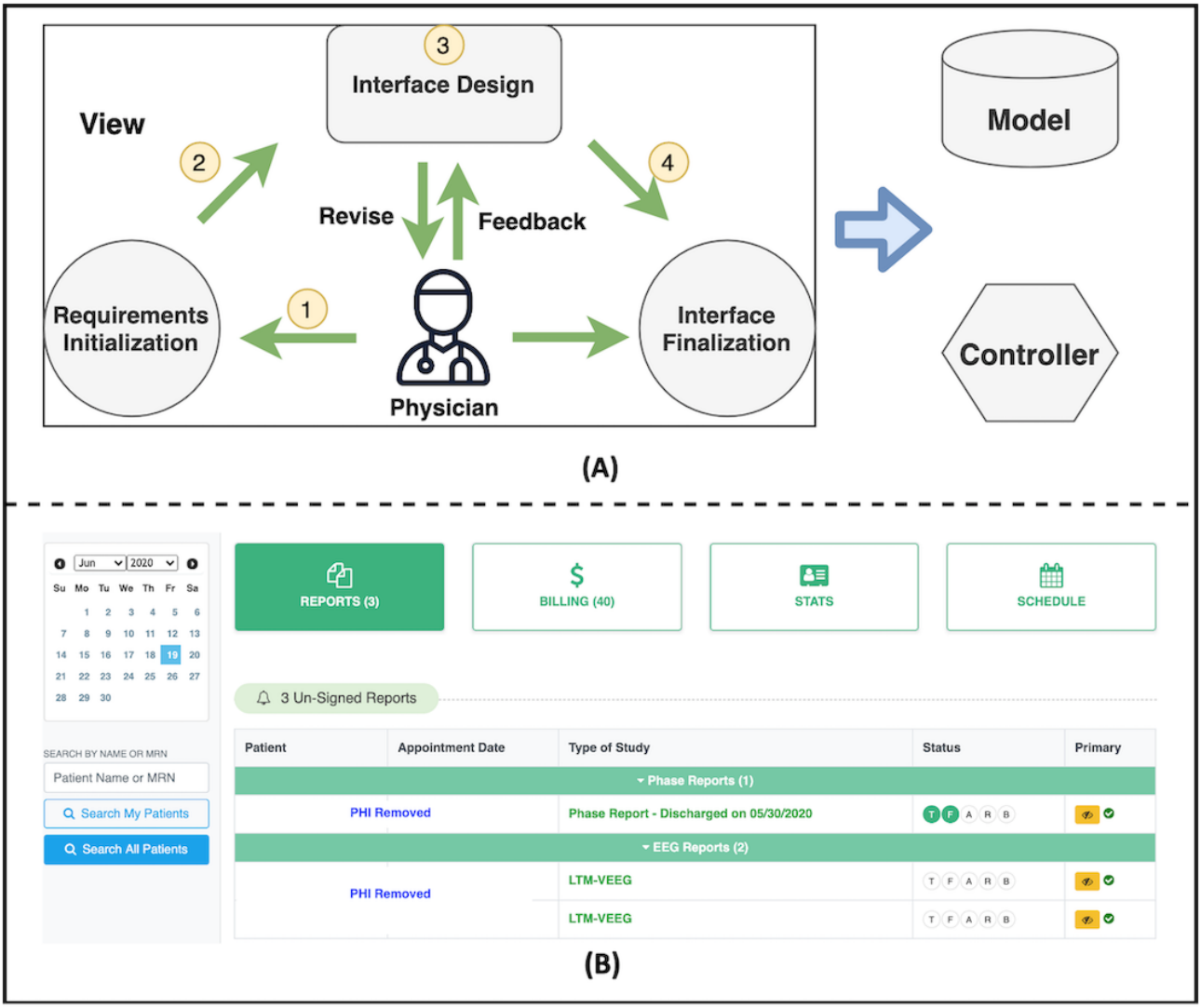
Physician interfaces play a critical role in EHR systems and are the most important factor affecting usability and clinical efficiency [12,29]. However, in the history of EHR development, physicians have rarely had a major role in deciding how an EHR interface should be built. Modern EHR systems (Allscripts, EPIC, and Cerner) offer physicians some opportunity to provide document templates, but physicians often neither have the expertise to optimize such templates nor do they have the flexibility to maintain or update these templates as needed.

In EpiToMe, we address this problem using an agile, physician-centered design and interface-driven development during all stages of the development process from inception. As depicted in Figure 1A, EpiToMe follows the classic model-view-controller architectural pattern. We use user interfaces to drive the development of data models and controllers. Our interface design process consists of 4 steps with physicians in the loop: (1) The process starts with physicians’ initial requirements; (2) then, informaticians complete the next

iteration of the interface prototype incorporating such requirements; (3) physicians give feedback about the prototype and working interfaces and suggest revisions to be made in the future iterations; and (4) step 3 continues iteratively until the design is accepted and finalized by physicians and a testing or production version is deployed.

With this physician-centered design and interface-driven development method, EpiToMe ensures that the interfaces have the look, feel, and functionality desired by physicians, improving user satisfaction and optimizing clinical efficiency (eg, Figure 1B).

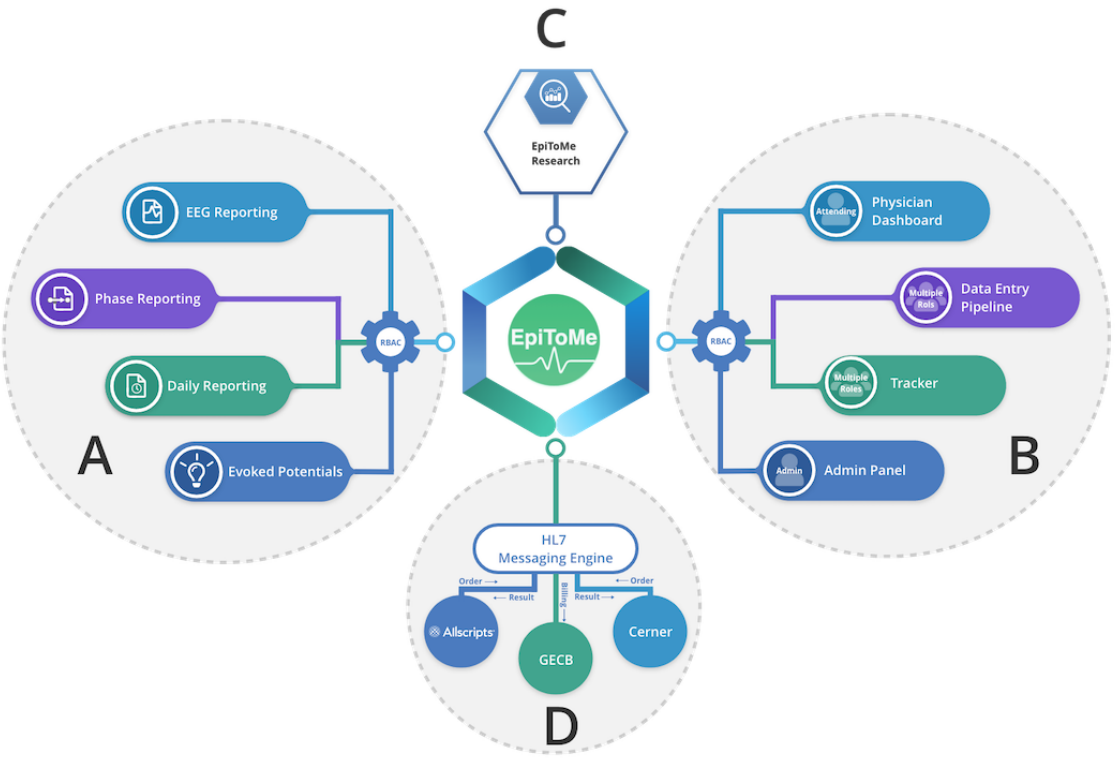
Figure 1. (A) Steps involved in our physician-centered design and interface-driven development process; (B) Exemplar physician dashboard resulting from our physician-centered design. LTM-VEEG: Long-term Video Electroencephalogram; PHI: Protected Health Information.



Functional Architecture

Figure 2 shows the functional architecture design of EpiToMe, consisting of 4 major areas. Figure 2A shows the data capture interfaces for clinical reports. Currently, 4 types of clinical reports are built in EpiToMe: EEG Reporting, Phase Reporting, Daily Reporting, and Evoked Potentials. These interfaces capture essential diagnostic information for epilepsy care, which are then seamlessly pushed to the general EHR (see Figure 2D). Figure 2B shows the data dashboards. The Physician Dashboard allows a physician to track outstanding reports, file for billing, monitor statistics of activity in a given time interval, and review the service schedule. The tracker is an interactive, real-time interface displaying each patient's status in the entire epilepsy care workflow, from admission and discharge from EMU to postoperative evaluations. Figure 2C shows the clinical research query interface. This is a faceted interface for ad-hoc, on-the-fly identification and construction of epilepsy patient subgroups for research. In EpiToMe, all patient information including demographics, diagnoses, epilepsy-related clinical characteristics, and medications are indexed to make such information queriable and exportable. Figure 2D shows the interoperability through an HL7 messaging engine. HL7 is a widely used protocol for the transfer of clinical and administrative data among EHR systems [36]. EpiToMe implemented 3 primary HL7 standard message types: orders, results, and charges for epilepsy care. The EpiToMe HL7 messaging engine allows it to seamlessly communicate with the general EHR. EpiToMe receives orders from the general EHR and sends back the completed reports and billing messages.

Figure 2. EpiToMe's functional architecture with 4 major functional areas: (A) data capture interfaces for clinical reports, (B) data and analytical dashboards, (C) clinical research query interface, (D) HL7 messaging engine for communication with background electronic health record systems (Allscripts, Cerner, and EPIC). Role-based access control can be configured to manage who gets which levels of access to what information, as defined by their clinical roles. EEG: electroencephalogram.



Collaborative, Asynchronous Care Documentation

Clinical care comprises collaborative teamwork from different clinical stakeholders; patient data documentation should follow suit. We designed the EpiToMe data entry pipeline to be automatically triggered when an order is placed in the general EHR. An order message containing patient demographics and order details is sent to EpiToMe. Based on the message, EpiToMe will create a new report document for the order and notify EEG technicians that a patient report is waiting to be handled. Next, an EEG technician will perform the EEG recording on patients and document the EEG specifications in EpiToMe. EpiToMe will mark the report as "technician completed" and pass it to clinical fellows who read the EEG recordings and enter their interpretations in the report. After fellows complete their data entries, the report will appear in the physician dashboard, and the clinical attending physicians will take it over, review it, and finalize it.

Role-Based Access Control for Collaborative Data Entry

EpiToMe applies a role-based access control (RBAC) method to manage users' access to data and interfaces. RBAC is a popular framework for implementing the security policy of an organization's enterprise information system. In RBAC, permissions are associated with roles, and roles are assigned to users. We designed EpiToMe's RBAC so that every user is assigned one or multiple roles, and each role defines what actions are allowed to perform within the system. To fully reflect the physician-centered interface design, RBAC needs to be implemented not only at the data access level but also at the interface level. Our RBAC method ensures that users can focus on the responsibility corresponding to individual clinical roles, thereby improving efficiency.

Ontology-Driven Data Capture

EpiToMe uses EpSO to provide a standard vocabulary and guide the data entry for all clinicians. EpSO provides more than 600 terms, which include epileptic diagnoses, epilepsy semiologies, epileptogenic zones, lateralizing signs, EEG activities, and etiologies. We designed a dedicated widget in the style of multilevel dropdowns for clinicians to enter patient data. This widget supports hierarchical "hover to expand" operation, allowing users to locate the desired terms efficiently. With the ontology-guided data entry method, users select data items instead of typing them, which prevents possible common data quality issues such as typos and inconsistencies.

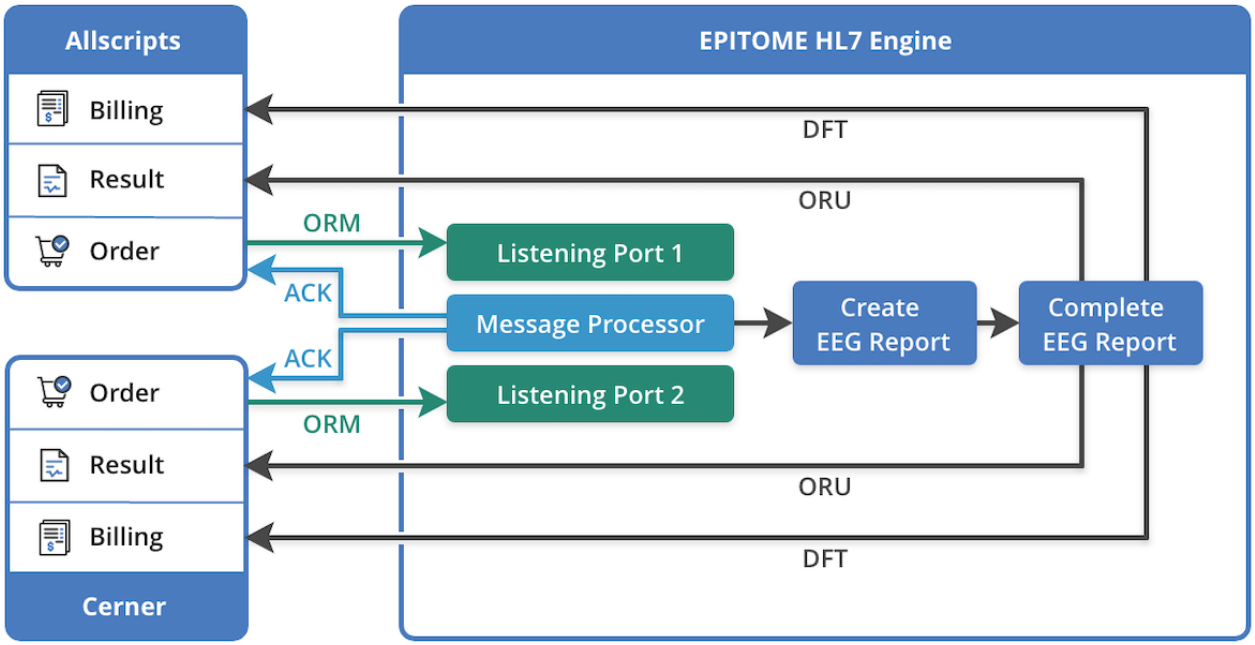
Interoperability Using HL7 Information Exchange

EpiToMe handles the epilepsy-related orders from multiple locations. Two different EHRs are used in these locations: Allscripts (EHR of University of Texas [UT] physicians) and Cerner (EHR of Memorial Hermann Health System). We designed an HL7 engine that can consume HL7 messages from multiple EHR vendors.

Figure 3 shows the architecture of the HL7 engine. HL7 messages are used to transfer electronic data between disparate health care systems. Each HL7 message sends information about a particular event such as a patient admission or a lab test order.

Three primary HL7 standard message types are handled in EpiToMe: orders (ORM), results (ORU), and charges (DFT). ORM messages contain patient demographic information and order-related data. ORU is usually in response to an order and provides clinical observations. DFT is used to send billing information. EpiToMe receives order messages from Allscripts and Cerner and creates patient reports with the embedded information in the order messages. EpiToMe confirms the order messages by sending back acknowledgement (ACK) messages. Physicians complete these reports in EpiToMe and then send them back to the original EHR system with ORU messages. After the results are accepted, physicians can continue to file billing messages using EpiToMe. Our HL7 engine design allows EpiToMe to seamlessly communicate with the parent EHR systems.

Figure 3. HL7 messaging interface between EpiToMe and the parent electronic health record systems. ACK: Acknowledgement; DFT: Detail Financial Transaction; EEG: electroencephalogram; ORM: Order Entry Message; ORU: Order Result.



Assessment Survey Design

To assess the EpiToMe's effectiveness in improving the user experience with patient documentation and reducing physician burnout, we designed an online survey administered within EpiToMe. Physicians users of the EpiToMe system were invited to participate in the survey. The survey (Table 1) consists of 10 questions addressing the common concerns about physicians' dissatisfaction with general EHR systems including the length of time spent on patient documentation, face-to-face interaction opportunities with patients, and catching up with care documentation in off-work time [13-16]. The specific EHR systems that are compared with EpiToMe in this survey are Allscripts and Cerner. Our design of the questionnaire was also informed by the Maslach Burnout Inventory Manual [37]. Questions 1-8 were designed to have Likert rating scales from 1 to 5, representing strongly disagree (1), disagree (2), neutral (3), agree (4), and strongly agree (5). Questions 9 and 10 are open ended to solicit input in free-text form. Specifically, question 9 asks physicians to enter which aspects of EpiToMe help with addressing physician burnout. Question 10 solicits the features physicians would like to see implemented in a future version of EpiToMe.

Table 1. The 10 survey questions and their answer options.

Question	Answer type
1. My overall workflow is less frustrating with EpiToMe compared to before.	1-5 rating scale
2. Completing patient reports with EpiToMe is easier and more intuitive than with EHR.	1-5 rating scale
3. I spend less amount of time billing using EpiToMe compared to using EHR.	1-5 rating scale
4. I spend less after-work time catching up with reports or billing using EpiToMe.	1-5 rating scale
5. EpiToMe allows me to spend more time on direct patient care.	1-5 rating scale
6. My oversight of the patient journey from the clinic to epilepsy surgery is better with EpiToMe.	1-5 rating scale
7. The dashboard in EpiToMe helps me to know my task list and complete it appropriately.	1-5 rating scale
8. For epilepsy reporting and billing, I would prefer using EpiToMe compared to EHR.	1-5 rating scale
9. In my opinion, the aspects of EpiToMe which help me address physician burnout are:	Free text
10. In my opinion, the additional features that I would like to see implemented in EpiToMe are:	Free text

Results

Physician Dashboard

Overview

With the physician-centered design, we created a physician dashboard—an integrated interface specially designed for epilepsy care providers. As shown in Figure 1B, the physician dashboard consists of 4 tabs: reports, billing, statistics, and schedule. EpiToMe directly leads physicians to this dashboard reflecting the present date status when they log in, where they can manage all the day-to-day tasks by selecting a specific day on the calendar without the need to navigate between different web pages or switch to different systems.

Reports

The default function tab is “Reports.” The number of outstanding reports to be completed is displayed in the bracket following the tab title. Physicians can review and complete reports here and send the completed reports back to the parent EHR systems with one button click.

Billing

After the reports are accepted by the parent EHR, physicians can continue to work on the billing. As a bespoke EHR system, EpiToMe automatically pulls all billing-related information and displays it in a user-friendly style in the billing interface. Physicians can file billing for a report with 3 to 4 clicks. In contrast, it takes more than 24 clicks on multiple pages to complete the same task in the billing interface of the general-purpose EHR system.

Statistics

The “Statistics” tab provides an overview of reports completed, documented, and billed by physicians, including the number of reports by month and type of study.

Schedule

In the “Schedule” tab, physicians can review their service schedules for the whole year. It allows a physician to send requests to another physician to switch schedules, which was a significant challenge in the previous schedule management system. The implementation of the schedule functionality also allows EpiToMe to automatically link reports to their corresponding physicians.

Usage Summary

EpiToMe creates interfaces for 4 types of reports for epilepsy care: EEG report, EMU phase report, EMU daily report, and evoked potentials. Table 2 shows the statistics for these reports. The EEG report is the first type of reporting function for production use in EpiToMe. By September 21, 2020, clinicians had completed 15,417 EEG reports in EpiToMe since its first launch on February 18, 2019. The EMU phase report is the second reporting interface in production use since July 1, 2019. A total of 1369 EMU phase reports have been completed in EpiToMe. EpiToMe also has documented 2635 EMU daily reports since its production date of November 15, 2019. The evoked potentials reporting function is under testing. Combining these reports, EpiToMe has documented 19,421 reports for 6593 unique patients.

Table 2. EpiToMe report statistics.

Report type	Status	Status date	Number of reports (N=19,421)	Number of patients (N=6593)
EEG[a] report	Production	02/18/2019	15,417	6382
EMU[b] phase report	Production	07/01/2019	1369	1053
EMU daily report	Production	11/15/2019	2635	312
Evoked potentials	Test	N/A[c]	0	0

[a]EEG: electroencephalogram.

[b]EMU: Epilepsy Monitoring Unit.

[c]N/A: not applicable.

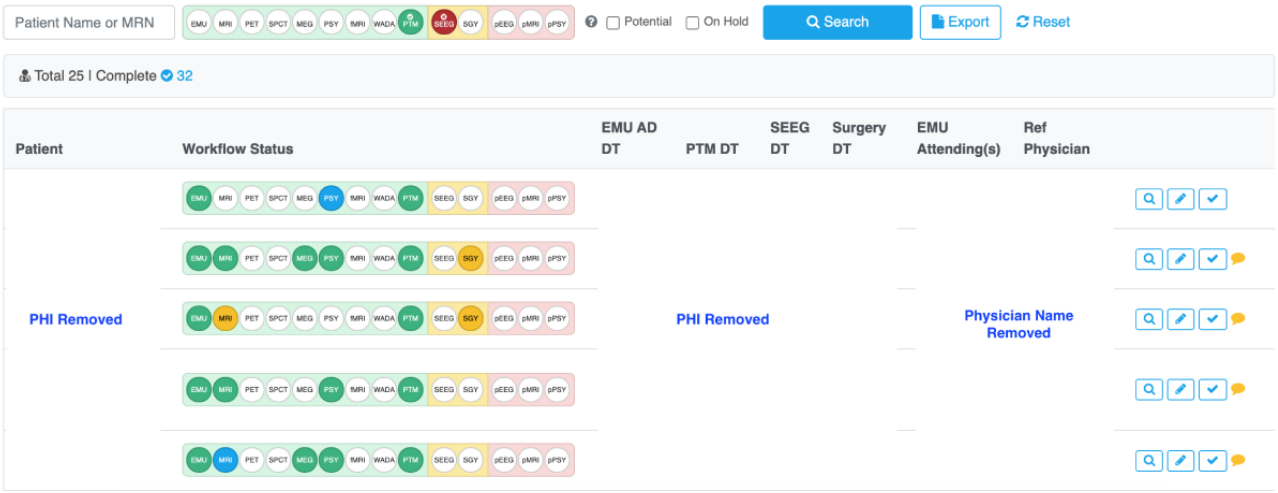
Patient Tracking Interface

Epilepsy care is a complex process that requires the collaboration of multiple clinical teams, including neurologists, radiologists, neuropsychologists, and neurosurgeons, especially for patients who are not responsive to seizure-control medications and treated as surgical candidates. It is critical to keep track of the patient status in the clinical workflow from 2 perspectives: (1) Keep different clinical teams on the same page, and (2) identify and resolve the bottleneck in the workflow. EpiToMe creates a patient tracking interface called a tracker. As displayed in Figure 4, the tracker records 14 possible steps of the patient journey in epilepsy care, which include admission to EMU, discharge from EMU, positron emission tomography scan, Ictal single-photon emission computed tomography, magnetoencephalography, neuropsychology, functional magnetic resonance imaging, Wada, patient management, stereoelectroencephalography, surgery, postoperative EEG, postoperative magnetic resonance imaging, and postoperative neuropsychology. Each circle represents a clinical step. Color codes are used to indicate the status of completion of each step: Blue indicates a test is ordered, yellow means a procedure is scheduled, and green shows that a step is complete.

The search functionality in the tracker provides a search mechanism for users to quickly find a patient or a cohort of patients. In addition to searching by name or medical record number, users can search patients by the completion status of each step and combine statuses to get results. In the search template, red status means a step has not started yet. Figure 4 shows an example of searching patients who have had patient management conferences but pending stereoelectroencephalography.

The tracker also provides comments functionality for users with the role of a nurse navigator, which is the specific user role responsible for updating the tracker in EpiToMe. The nurse navigator will enter comments once a bottleneck is identified and notify all related clinical teams to keep them alert and encourage teamwork to resolve the bottleneck.

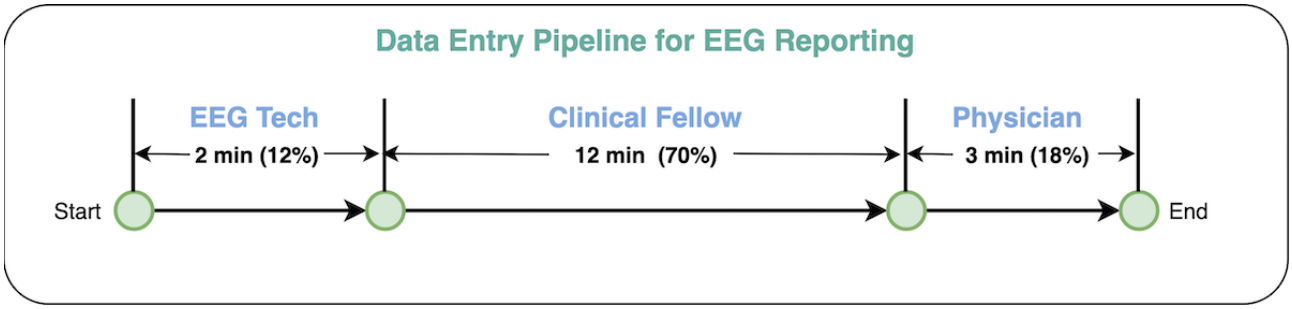
Figure 4. Tracker of epilepsy surgery candidates. AD: admission; DT: date; EMU: Epilepsy Monitoring Unit; MRN: medical record number; PHI: protected health information; PTM: patient management; Ref: referring; SEEG: stereoelectroencephalography.



Collaborative, Asynchronous, Data Entry Pipeline

With the physician-desired interface and RBAC in place, EpiToMe implements a collaborative data entry pipe to improve clinical efficiency and distribute patient documentation workload. Figure 5 shows the usage statistics of the data entry pipeline with EEG reporting. In this pipeline, the physician is not the only role that completes patient documentation. Instead, the result of inspecting the user activity logs shows that the physician's average time spent on the EEG reporting only occupies 18% of the total time of all clinical roles.

Figure 5. Collaborative data entry pipeline which reduces physician burden with patient documentation. EEG: electroencephalogram.



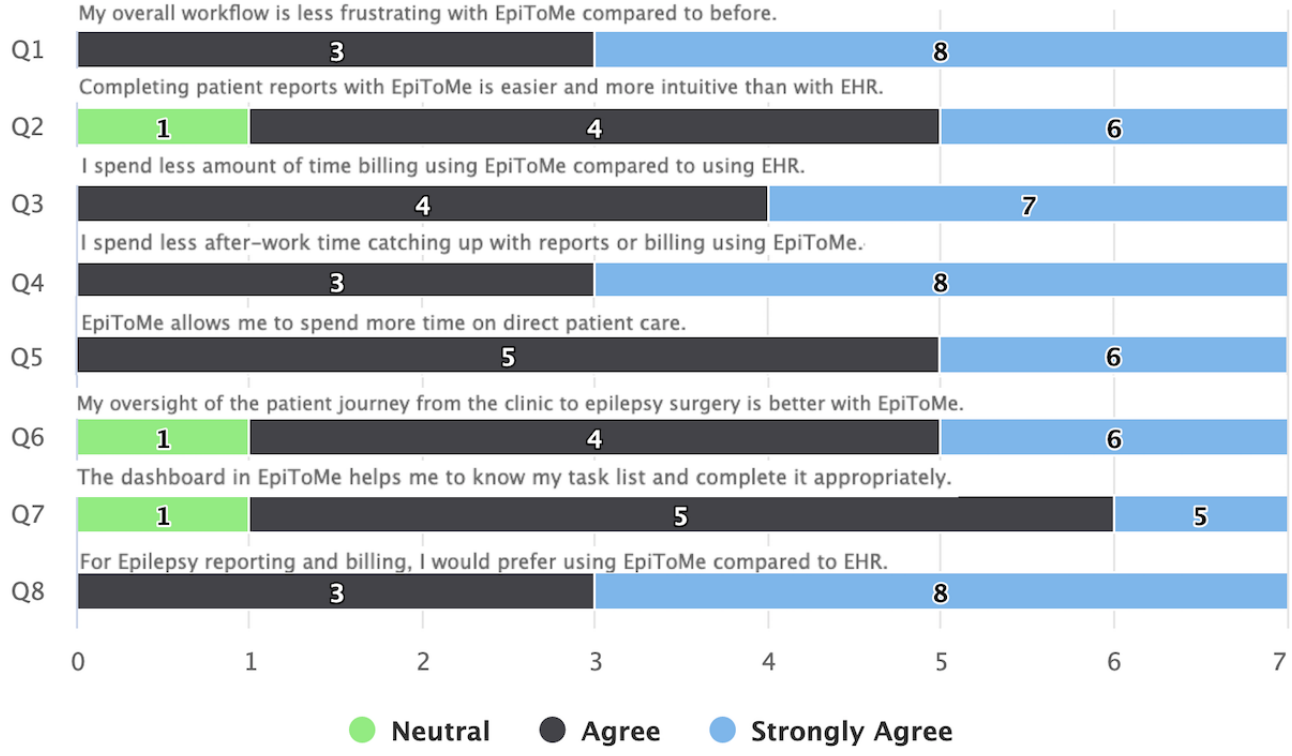
Survey Results

Of the 16 survey invitations sent to physicians, we received 11 completed responses. Answers to each of the questions 1-8 are displayed in Figure 6. In general, physicians favored EpiToMe for patient reporting and billing compared to the general EHR. All physicians agreed that the workflow with EpiToMe was less frustrating, and most (8/11, 73%) strongly agreed on this. Most physicians (10/11, 91%) thought EpiToMe made it easier to complete patient reports, and all physicians agreed that billing with EpiToMe took less time than previously using the general EHR systems. As a result, all physicians reported in question 4 that they spent less off-work time catching up with reports or billing, which is considered a significant factor for physician burnout. All physicians agreed that EpiToMe allowed them to spend more time on direct patient care, and most physicians (10/11, 91%) thought EpiToMe helped them better manage the patient journey and their task list.

For question 9, physicians described the aspects of EpiToMe that helped them address physician burnout. One respondent said: “Intuitive, fast, accurate, comprehensive interactions; everything makes sense; very little redundancy.” Another respondent pointed out “ease of reporting and billing” as well as “no need of separate data sets (eg, personal spreadsheets, email lists) to keep track of patients and testing.” Another physician endorsed the integrating role of EpiToMe as “It reduces the number of places (different EHR systems) I need to be working on simultaneously—the integration between reporting and billing reduces the amount of time spent on non-patient–care related work.” One physician also appreciated the collaborative data entry pipeline: “better time management and easier/more efficient interface with the fellows in terms of report writing and billing on time.”

In response to question 10, physicians proposed many constructive suggestions that can serve as future directions for EpiToMe. These include “more EHR interfaces,” “outpatient module,” and being more user-friendly such as “being able to have multiple sections within the report open while editing the report.”

Figure 6. Responses for survey questions 1-8. EHR: electronic health records.



Discussion

Physician Feedback

With a specialty-specific, physician-centered interface design, EpiToMe can improve overall clinical efficiency. For example, for EEG reporting, the average time delay (from the completion of EEG recording to the finalization of the EEG reporting) to complete an EEG report is 14 hours and 30 minutes for 98% of EEG reports. Patients obtain their EEG reports within the same day of visit rather than a couple of days or weeks later using the previous general-purpose EHR. Survey feedback by clinical attending physicians showed significant preference for using EpiToMe to perform reporting and billing tasks compared with general-purpose EHR systems. Within the 88 answers from 11 senior clinical attending physicians, 85 (96.6%) indicated that EpiToMe is better than the general EHR on specific tasks, with only 3 neutral answers.

Multisite Deployment

EpiToMe is a multisite system supporting interoperability with multiple types of EHR systems. Currently, EpiToMe has been deployed at 4 clinical centers including UT Physicians Clinic, Memorial Hermann Texas Medical Center, Memorial Hermann The Woodlands Medical Center, and Memorial Hermann Cypress Hospital. Within these medical centers, UT Physicians Clinic uses Allscripts as their general-purpose EHR system, while others use Cerner. We are making EpiToMe also interoperable with Epic as UT Physicians Clinic transitions from Allscripts to Epic.

Interoperability

HL7 is a widely used standard for data exchange in clinical information systems. Implementation of an HL7 messaging engine enables EpiToMe to interoperate and exchange information with general-purpose EHR systems, resulting in significantly reduced or eliminated redundant data entry. Our HL7 messaging engine also makes EpiToMe scalable: New clinical centers can be added in EpiToMe as long as their EHR platforms offer service to support HL7 communication standards.

Generalizability

EpiToMe is not a replacement for general-purpose EHR. Instead, it is complementary to existing general EHR solutions. EpiToMe relies on the availability of the parent EHR to admit and transmit patient demographic information and epilepsy-related orders, which triggers the corresponding, additional data capture process in EpiToMe. Although EpiToMe is optimized for epilepsy care, our methodology, design principle, system architecture, and interface elements are generally applicable. In fact, we are applying a similar approach in UTHealth Neurosciences service lines to derive similar benefits for other clinical specialties.

Survey Limitations

Our physician survey study is preliminary, as this is not the primary focus of the paper. Limitations of this survey include the small sample of survey participants and the lack of complete independence of survey participants and the informatics development team. Such limitations may present hidden bias in survey results, and larger-scale, anonymous surveys are the preferred approach for feedback. However, production-level deployment and long-term operation of interoperable bespoke EHRs implemented using physician-centered design and the latest information technology for clinical specialties are uncommon. Therefore, timely assessment of such bespoke EHRs, even on a smaller scale and with limitations, provides valuable and much-needed operational feedback to inform hospitals' adoption strategies. Surveys should also accommodate a strategy tailored to the tremendous effort and longer development cycle needed in designing, deploying, and operationalizing such systems in real-world clinical practice and patient care settings. After a few years of operation, a more systematic, comparative activity log analysis would provide more objective insight about where our bespoke approach made the most impact and where further enhancements may be needed.

Conclusions

Working closely with physicians, we used an interface-driven development approach to create EpiToMe de novo, to embody physician preferences and optimize clinical workflow for epilepsy care while ensuring interoperability with the parent EHR. EpiToMe offers an exemplar pathway to mitigate physician burnout and improve the quality and productivity of care by combining physician-centered design with the latest advances in information technology in a bespoke EHR system.

Acknowledgments

This research was supported by the National Institutes of Health (NIH) through grants U01NS090408, U01NS090405, and R01NS116287. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

We thank Wei-Chun Chou for helping with drawing the EpiToMe's system architecture graph and enhancing the user interfaces.

Authors' Contributions

GQZ conceptualized this study. ST implemented the system with contributions from GQZ, SL, LC, and JH. SL, GQZ, ST, JH, and LC designed the survey. ST and SL administrated the survey. ST and GQZ developed and refined the manuscript with contributions from JH, LC, and SL.

Conflicts of Interest

None declared.

References

1. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471 [FREE Full text] [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
2. King J, Patel V, Jamoom EW, Furukawa MF. Clinical benefits of electronic health record use: national findings. Health Serv Res 2014 Feb;49(1 Pt 2):392-404 [FREE Full text] [doi: 10.1111/1475-6773.12135] [Medline: 24359580]
3. Zlabek JA, Wickus JW, Mathiason MA. Early cost and safety benefits of an inpatient electronic health record. J Am Med Inform Assoc 2011;18(2):169-172 [FREE Full text] [doi: 10.1136/jamia.2010.007229] [Medline: 21292703]
4. Blumenthal D. Launching HITECH. N Engl J Med 2010 Feb 04;362(5):382-385. [doi: 10.1056/NEJMp0912825] [Medline: 20042745]
5. Jha AK, DesRoches CM, Campbell EG, Donelan K, Rao SR, Ferris TG, et al. Use of electronic health records in U.S. hospitals. N Engl J Med 2009 Apr 16;360(16):1628-1638. [doi: 10.1056/NEJMsa0900592] [Medline: 19321858]
6. Jones SS, Adams JL, Schneider EC, Ringel JS, McGlynn EA. Electronic health record adoption and quality improvement in US hospitals. Am J Manag Care 2010 Dec;16(12 Suppl HIT):SP64-SP71 [FREE Full text] [Medline: 21314225]
7. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital. J Am Med Inform Assoc 2017 Nov 01;24(6):1142-1148 [FREE Full text] [doi: 10.1093/jamia/ocx080] [Medline: 29016973]
8. Thakkar M, Davis DC. Risks, barriers, and benefits of EHR systems: a comparative study based on size of hospital. Perspect Health Inf Manag 2006 Aug 14;3:5 [FREE Full text] [Medline: 18066363]
9. Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. Risk Manag Healthc Policy 2011;4:47-55 [FREE Full text] [doi: 10.2147/RMHP.S12985] [Medline: 22312227]
10. Walker J, Pan E, Johnston D, Adler-Milstein J, Bates DW, Middleton B. The value of health care information exchange and interoperability. Health Aff (Millwood) 2005 Jan;Suppl Web Exclusives:W5-10. [doi: 10.1377/hlthaff.w5.10] [Medline: 15659453]
11. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752 [FREE Full text] [doi: 10.7326/0003-4819-144-10-200605160-00125] [Medline: 16702590]
12. Grams R. The new America Electronic Medical Record (EMR) design criteria and challenge. J Med Syst 2009 Dec;33(6):409-411. [doi: 10.1007/s10916-009-9319-0] [Medline: 20052893]
13. Tai-Seale M, Olson CW, Li J, Chan AS, Morikawa C, Durbin M, et al. Electronic Health Record Logs Indicate That Physicians Split Time Evenly Between Seeing Patients And Desktop Medicine. Health Aff (Millwood) 2017 Apr 01;36(4):655-662 [FREE Full text] [doi: 10.1377/hlthaff.2016.0811] [Medline: 28373331]
14. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan W, Sinsky CA, et al. Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations. Ann Fam Med 2017 Sep;15(5):419-426 [FREE Full text] [doi: 10.1370/afm.2121] [Medline: 28893811]
15. Collier R. Electronic health records contributing to physician burnout. CMAJ 2017 Nov 13;189(45):E1405-E1406 [FREE Full text] [doi: 10.1503/cmaj.109-5522] [Medline: 29133547]
16. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. J Intern Med 2018 Jun;283(6):516-529 [FREE Full text] [doi: 10.1111/joim.12752] [Medline: 29505159]
17. Rothenberger DA. Physician Burnout and Well-Being: A Systematic Review and Framework for Action. Dis Colon Rectum 2017 Jun;60(6):567-576. [doi: 10.1097/DCR.0000000000000844] [Medline: 28481850]
18. Lacy BE, Chan JL. Physician Burnout: The Hidden Health Care Crisis. Clin Gastroenterol Hepatol 2018 Mar;16(3):311-317. [doi: 10.1016/j.cgh.2017.06.043] [Medline: 28669661]
19. Sigsbee B, Bernat JL. Physician burnout: A neurologic crisis. Neurology 2014 Dec 09;83(24):2302-2306. [doi: 10.1212/WNL.0000000000001077] [Medline: 25378679]
20. Dzau VJ, Kirch DG, Nasca TJ. To Care Is Human - Collectively Confronting the Clinician-Burnout Crisis. N Engl J Med 2018 Jan 25;378(4):312-314. [doi: 10.1056/NEJMp1715127] [Medline: 29365296]
21. Jackson TN, Pearcy CP, Khorgami Z, Agrawal V, Taubman KE, Truitt MS. The Physician Attrition Crisis: A Cross-Sectional Survey of the Risk Factors for Reduced Job Satisfaction Among US Surgeons. World J Surg 2018 May;42(5):1285-1292. [doi: 10.1007/s00268-017-4286-y] [Medline: 29067517]
22. Williams ES, Manwell LB, Konrad TR, Linzer M. The relationship of organizational culture, stress, satisfaction, and burnout with physician-reported error and suboptimal patient care: results from the MEMO study. Health Care Manage Rev 2007;32(3):203-212. [doi: 10.1097/01.HMR.0000281626.28363.59] [Medline: 17666991]
23. Olson KD. Physician Burnout-A Leading Indicator of Health System Performance? Mayo Clin Proc 2017 Nov;92(11):1608-1611. [doi: 10.1016/j.mayocp.2017.09.008] [Medline: 29101929]

24. Kushnir T, Greenberg D, Madjar N, Hadari I, Yermiahu Y, Bachner YG. Is burnout associated with referral rates among primary care physicians in community clinics? Fam Pract 2014 Feb;31(1):44-50 [FREE Full text] [doi: 10.1093/fampra/cmt060] [Medline: 24148815]
25. Sinsky CA, Dyrbye LN, West CP, Satele D, Tutty M, Shanafelt TD. Professional Satisfaction and the Career Plans of US Physicians. Mayo Clin Proc 2017 Nov;92(11):1625-1635. [doi: 10.1016/j.mayocp.2017.08.017] [Medline: 29101932]
26. Gardner RL, Cooper E, Haskell J, Harris DA, Poplau S, Kroth PJ, et al. Physician stress and burnout: the impact of health information technology. J Am Med Inform Assoc 2019 Feb 01;26(2):106-114 [FREE Full text] [doi: 10.1093/jamia/ocy145] [Medline: 30517663]
27. Khairat S, Coleman C, Ottmar P, Jayachander DI, Bice T, Carson SS. Association of Electronic Health Record Use With Physician Fatigue and Efficiency. JAMA Netw Open 2020 Jun 01;3(6):e207385 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.7385] [Medline: 32515799]
28. Redd TK, Doberne JW, Lattin D, Yackel TR, Eriksson CO, Mohan V, et al. Variability in Electronic Health Record Usage and Perceptions among Specialty vs. Primary Care Physicians. AMIA Annu Symp Proc 2015;2015:2053-2062 [FREE Full text] [Medline: 26958305]
29. Butterfield B. What physicians want in an EHR: physician practices have been ignored by health IT vendors until now. J AHIMA 2012;83(11):44-6; quiz 47. [Medline: 23210297]
30. Sieja A, Markley K, Pell J, Gonzalez C, Redig B, Kneeland P, et al. Optimization Sprints: Improving Clinician Satisfaction and Teamwork by Rapidly Reducing Electronic Health Record Burden. Mayo Clin Proc 2019 May;94(5):793-802 [FREE Full text] [doi: 10.1016/j.mayocp.2018.08.036] [Medline: 30824281]
31. Sahoo SS, Zhao M, Luo L, Bozorgi A, Gupta D, Lhatoo SD, et al. OPIC: Ontology-driven Patient Information Capturing system for epilepsy. AMIA Annu Symp Proc 2012;2012:799-808 [FREE Full text] [Medline: 23304354]
32. Zhang GQ, Cui L, Lhatoo S, Schuele SU, Sahoo SS. MEDCIS: Multi-Modality Epilepsy Data Capture and Integration System. AMIA Annu Symp Proc 2014;2014:1248-1257 [FREE Full text] [Medline: 25954436]
33. Tao S, Walter BL, Gu S, Zhang GQ. Web-interface-driven development for Neuro3D, a clinical data capture and decision support system for deep brain stimulation. In: Yin X, Geller J, Li Y, Zhou R, Wang H, Zhang Y, editors. Health Information Science. HIS 2016. Lecture Notes in Computer Science, vol 10038. Cham, Switzerland: Springer Publishing Company; 2016:31-42.
34. Tao S, Cui L, Wu X, Zhang GQ. Facilitating Cohort Discovery by Enhancing Ontology Exploration, Query Management and Query Sharing for Large Clinical Data Repositories. AMIA Annu Symp Proc 2017;2017:1685-1694 [FREE Full text] [Medline: 29854239]
35. Sahoo SS, Lhatoo SD, Gupta DK, Cui L, Zhao M, Jayapandian C, et al. Epilepsy and seizure ontology: towards an epilepsy informatics infrastructure for clinical research and patient care. J Am Med Inform Assoc 2014;21(1):82-89 [FREE Full text] [doi: 10.1136/amiajnl-2013-001696] [Medline: 23686934]
36. Health Level7. HL7 International. 2021. URL: http://www.hl7.org [accessed 2020-09-19]
37. Kleijweg JHM, Verbraak MJPM, Van Dijk MK. The clinical utility of the Maslach Burnout Inventory in a clinical population. Psychol Assess 2013 Jun;25(2):435-441. [doi: 10.1037/a0031334] [Medline: 23356679]

Abbreviations

EEG: electroencephalogram
EHR: electronic health records
EMU: Epilepsy Monitoring Unit
EpSO: epilepsy and seizure ontology
HITECH: Health Information Technology for Economic and Clinical Health
RBAC: role-based access control
UT: University of Texas

Edited by G Eysenbach, R Kukafka; submitted 28.07.20; peer-reviewed by S Lin, R Marshall, M Zanetti, C Xu; comments to author 12.09.20; revised version received 21.09.20; accepted 17.12.20; published 12.02.21.

Please cite as:
Tao S, Lhatoo S, Hampson J, Cui L, Zhang GQ
A Bespoke Electronic Health Record for Epilepsy Care (EpiToMe): Development and Qualitative Evaluation
J Med Internet Res 2021;23(2):e22939
URL: http://www.jmir.org/2021/2/e22939/
doi:10.2196/22939
PMID:33576745

Original Paper

Effect of Sleep and Biobehavioral Patterns on Multidimensional Cognitive Performance: Longitudinal, In-the-Wild Study

Manasa Kalanadhabhatta[1], BTech; Tauhidur Rahman[1], PhD; Deepak Ganesan[1], PhD

College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, MA, United States

Corresponding Author:
Manasa Kalanadhabhatta, BTech
College of Information and Computer Sciences
University of Massachusetts Amherst
140 Governors Drive
Amherst, MA, 01003
United States
Phone: 1 4135453819
Email: manasak@cs.umass.edu

Abstract

Background: With nearly 20% of the US adult population using fitness trackers, there is an increasing focus on how physiological data from these devices can provide actionable insights about workplace performance. However, in-the-wild studies that understand how these metrics correlate with cognitive performance measures across a diverse population are lacking, and claims made by device manufacturers are vague. While there has been extensive research leading to a variety of theories on how physiological measures affect cognitive performance, virtually all such studies have been conducted in highly controlled settings and their validity in the real world is poorly understood.

Objective: We seek to bridge this gap by evaluating prevailing theories on the effects of a variety of sleep, activity, and heart rate parameters on cognitive performance against data collected in real-world settings.

Methods: We used a Fitbit Charge 3 and a smartphone app to collect different physiological and neurobehavioral task data, respectively, as part of our 6-week-long in-the-wild study. We collected data from 24 participants across multiple population groups (shift workers, regular workers, and graduate students) on different performance measures (vigilant attention and cognitive throughput). Simultaneously, we used a fitness tracker to unobtrusively obtain physiological measures that could influence these performance measures, including over 900 nights of sleep and over 1 million minutes of heart rate and physical activity metrics. We performed a repeated measures correlation (r_{rm}) analysis to investigate which sleep and physiological markers show association with each performance measure. We also report how our findings relate to existing theories and previous observations from controlled studies.

Results: Daytime alertness was found to be significantly correlated with total sleep duration on the previous night (r_{rm}=0.17, P<.001) as well as the duration of rapid eye movement (r_{rm}=0.12, P<.001) and light sleep (r_{rm}=0.15, P<.001). Cognitive throughput, by contrast, was not found to be significantly correlated with sleep duration but with sleep timing—a circadian phase shift toward a later sleep time corresponded with lower cognitive throughput on the following day (r_{rm}=–0.13, P<.001). Both measures show circadian variations, but only alertness showed a decline (r_{rm}=–0.1, P<.001) as a result of homeostatic pressure. Both heart rate and physical activity correlate positively with alertness as well as cognitive throughput.

Conclusions: Our findings reveal that there are significant differences in terms of which sleep-related physiological metrics influence each of the 2 performance measures. This makes the case for more targeted in-the-wild studies investigating how physiological measures from self-tracking data influence, or can be used to predict, specific aspects of cognitive performance.

(J Med Internet Res 2021;23(2):e23936) doi:10.2196/23936

KEYWORDS

fitness trackers; cognitive performance; alertness; cognitive throughput; sleep; activity; circadian rhythms

Introduction

Background

The cognitive functioning of an individual, characterized by a range of neurobehavioral metrics such as alertness, working memory, and cognitive throughput, is subject to systematic interday and intraday fluctuations [1]. These fluctuations are driven by the interplay of 3 biological processes: a circadian component *C*, a homeostatic component *H*, and a sleep inertia component *W* [2]. The circadian process *C*, governed by the circadian pacemaker (commonly referred to as the "body clock") located in the hypothalamus, is an endogenous oscillatory process with a roughly 24-hour period. The exact phase of the circadian pacemaker in each individual is subject to variation and is determined by their chronotype—a measure of their early-bird or late-owlness [3]. This measure determines their individual time of minimum sleep propensity or peak cognitive performance during the day. Sleeping out of phase with one's circadian rhythm—which is especially common among, but not limited to, populations such as shift workers [4]—leads to a phenomenon known as circadian misalignment. This misalignment has been shown to impact work-related fatigue [5], academic and work performance [6,7], performance on memory tests [8], etc.

In addition to this endogenous circadian rhythm, prior work has also established that an increase in the duration of wakefulness contributes to an exponential increase in sleep propensity, or homeostatic pressure [9]. This rise in homeostatic pressure *H* results in a corresponding decline in alertness and cognitive performance [10]. Other physiological measures such as heart rate variability (HRV) [11,12] and physical activity [13] have also been investigated to determine their impact on cognitive function [14]. Cognitive performance has also been found to be influenced by a variety of other factors, including sleep patterns [15], exposure to light [16], and consumption of caffeine [17].

Such fluctuations in cognitive performance are known to affect an individual's productivity on the job and influence work-related fatigue levels [5], as well as increase the risk of occupational hazards and accidents [18]. Therefore, there is considerable interest in understanding the nature of these fluctuations in real-world settings. However, much of prior work has been done within strict experimental protocols to study the effect of individual variables while controlling for all other factors [19,20]. Thus, the effect of multiple factors on cognitive performance in real-world settings is not as well-understood.

In this work, we attempt to discern the effect of various physiological measures on cognitive performance in a real-world setting and contextualize our findings within existing theories proposed based on controlled experiments.

Performance Prediction and Intervention Strategies

Given the widespread use of wearable devices and fitness trackers, recent efforts have explored using physiological data from these devices to characterize in-the-wild cognitive fluctuations. For example, one of the focus areas of Alphabet's Moonshot Factory is developing a "daytime score" that goes beyond yielding data from the previous night's sleep by harnessing this datum to tell users how prepared they are for the next day [21]. Various commercial fitness trackers are also advertised as providing metrics that can help track workplace productivity and cognitive performance [22,23].

There has been increasing work in the mobile health (mHealth) community on evaluating whether performance measures can be predicted from sleep- and circadian rhythm–based features. For instance, Abdullah et al [24] showed that alertness can be measured in the wild using a smartphone-based version of the Psychomotor Vigilance Task (PVT) [25]. Mark et al [26] studied engagement in the workplace and reported how job-related stress levels depend on the type of work being performed, and how rhythms of attention states can be identified in the workplace environment. Prior work has also looked into the effect of sleep duration and sleep debt on productivity and the use of technology among students [27]. Wahl and Amft [28] explored chronotype estimation using smartphone usage information and the 2-process circadian/homeostatic model of sleep regulation [9]. Althoff et al [29] used the speed of keystroke and click interactions on a web search engine as measures of cognitive performance and show that both metrics follow a circadian trend. Murnane et al [30] attempted to estimate sleep timings and circadian disruption from technology-mediated social interactions and, in a later study, showed that coarse-grained alertness levels can be gauged from app use features [31]. Abdullah et al [24] used sleep metrics estimated from smartphone usage to predict performance on the PVT test.

However, most studies to date have mainly focused on a single indicator of cognitive performance (eg, [24,29,31]). We argue that it is imperative to consider multiple dimensions of performance separately, because different aspects of cognitive function (eg, alertness, working memory, decision making) are known to be differently affected by factors such as sleep loss [32]. Therefore, to fully understand the effects of physiological variables on cognitive performance, one needs to take into account the multiplicity of these measures. This is especially important because different tasks may require different cognitive capabilities in order to be completed safely and responsibly. Further, many studies are constrained to a specific population, such as students or office workers, while studying cognitive performance. Our work targets a diverse population with an aim to generalize our understanding of the impact of physiological variables on 2 different cognitive measures.

Objective of the Study

While there is substantial work on studying variations in performance, there are 3 key gaps that we try to bridge in this work: (1) connecting in-the-wild studies to theories on the influence of sleep, activity, and circadian/homeostatic rhythms on performance; (2) taking into account multiple performance measures; and (3) generalizing our analysis across diverse populations.

We examine the effects of multiple physiological parameters on different dimensions of performance in a real-world setting, and thereafter contextualize our observations in the space of existing theories. To this end, our work investigates performance

across 2 axes: alertness and cognitive throughput. Through a 6-week-long research study, we collect physiological data relating to sleep and physical activity using a Fitbit fitness tracker and neurobehavioral task performance measures through a dedicated Android app. We deliberately ensure that we consider a diverse pool of participants with varying sleep and work patterns, recruiting individuals part of a regular workforce, shift workers, and graduate students. We then investigate the effects of a range of physiological and sleep-related parameters on alertness and cognitive throughput.

Our results provide insights into how sleep and activity metrics relate to work performance in different ways across alertness/cognitive throughput performance measures. We show that while alertness is sensitive to sleep duration as well as sleep stages, cognitive throughput exhibits no deteriorating effect from lack of sleep. By contrast, irregular sleep timings (ie, earlier or later bedtimes and wake-up times) indicate a significant effect on cognitive throughput, but not on alertness. We also quantify the influence of time of day, heart rate, and physical activity on both these measures.

Methods

Participants

We recruited 24 participants (14 female and 10 male), including 8 graduate students, 7 regular workers, and 9 shift workers, to participate in a 6-week-long research study between February and December 2019. These groups were chosen in order to ensure that our study encompasses a diverse set of participants with varying sleep and work patterns. Regular workers were individuals working jobs with a 9-to-5 schedule (or similar 8-hour daytime working hours) from Mondays to Fridays. Graduate students loosely followed a 9-to-5 weekday schedule, but reported that they also frequently worked late, occasionally worked on weekends, and had at least some flexibility in choosing their work hours. Shift workers in our study, by contrast, worked varying hours on different days of the week with shifts ranging from 8 hours to 24 hours in duration.

The study participants were recruited through convenience sampling roughly stratified by the groups mentioned above. The study was publicized through emailing lists at the authors' institution as well as flyers posted in the surrounding area. The local police and fire departments were also contacted in an attempt to recruit emergency responders. Three of the shift workers in the study were firefighters and 1 was a police dispatcher, 3 other shift workers worked in the service industry, and 1 worked as a transit driver. This gave us the unique opportunity to study patterns of alertness and cognitive throughput among emergency responders and essential workers. All participants were between 20 and 42 years of age (mean age 28 years). One graduate student had previously been treated for insomnia and 1 shift worker had been previously diagnosed with attention-deficit/hyperactivity disorder. None of the other participants reported any history of being diagnosed with either sleep or cognitive disorders.

Study Protocol

We used a Fitbit Charge 3 (Fitbit Inc.) and a smartphone app to collect different physiological and neurobehavioral task data, respectively, as part of our 6-week-long in-the-wild study.

Smartphone-Based Alertness and Cognitive Throughput Measures

Prior studies such as [32] have found that sleep deprivation can have varying effects on tasks that require sustained information processing as opposed to vigilance-based tasks. Therefore, we elected to study 2 fundamentally different aspects of human performance: alertness and cognitive throughput. We collected these metrics using an Android app to measure in-the-wild cognitive performance that was based on the toolkit produced by Dingler et al [33].

The app includes a questionnaire asking participants to rate their subjective sleepiness levels on the Karolinska Sleepiness Scale [34], and whether they had consumed a caffeinated drink in the last hour. This is followed by a task battery comprising the Psychomotor Vigilance Test (PVT; [25]) to measure alertness and an Addition Test (ADD; [35,36]) to measure cognitive throughput.

The PVT is a standard tool used to measure momentary alertness consisting of a 10-minute reaction time test that presents test takers with a visual stimulus at random intervals. The user has to press a button in response to these stimuli, and the response time is used as an objective marker of momentary alertness. Previous research has also validated the use of shorter, smartphone-based PVT tests [37-39], prompting us to use a 2-minute version of the test administered through the Android app.

Similar to prior studies such as [4,24], we studied fluctuations in alertness in terms of percentage deviation from the individual's response time at baseline, or relative response time. Because each PVT task session consists of multiple stimuli that the participants respond to, we first compute the median response time $MRT_{s,p}$ for each session s completed by participant p. In doing so, we ignore all instances of false clicks where the participant taps the screen before the stimulus appears. We then discard all sessions where $MRT_{s,p}$ exceeds 800 ms (in comparison, the standard threshold to classify a reaction as a lapse is 500 ms [40]). Further, we remove sessions where $MRT_{s,p}$ falls outside 3SDs of the mean as outliers, and calculate $(mean\_MRT)_p$ for each participant as the average $MRT_{s,p}$ across all sessions completed by p. Then, we calculate $RRT_{s,p}$ for each of their sessions as

$$RRT_{s,p} = \{1 - ([MRT_{s,p}]/[mean\_MRT]_p)\} \times 100$$

where RRT is relative response time.

It is important to note that while higher values of response time indicate lower alertness, higher values of relative response time indicate increased alertness.

Cognitive throughput is measured using a 1-minute addition/calculation performance test (ie, ADD) [35,36], where participants sum as many pairs of 2-digit numbers as possible within a fixed duration. The user's cognitive performance is

calculated as a percentage deviation from the user's baseline, or the relative number of additions attempted. Similar to the calculation of relative response time above, we first find $NAA_{s,p}$, that is, the number of additions attempted by participant p in session s. We then remove sessions where $NAA_{s,p}$ lies more than 3 SDs from the mean across all sessions for that participant. We further calculate $(mean\_NAA)_p$ for each participant across all sessions, and compute $RAA_{s,p}$ as

$$RAA_{s,p} = \{([NAA_{s,p}]/[mean\_NAA]_p) - 1\} \times 100$$

where RAA is relative number of additions attempted.

As with relative response times, a higher RAA indicates higher cognitive throughput.

Figure 1 shows the sleepiness and caffeination questionnaire in our Android app, while Figures 2 and 3 show the PVT and ADD tasks that participants were asked to complete. Throughout the 6-week-long study period, participants were asked to complete the task battery at least four times a day, with at least two hours between each session. As a reminder, the study app issued push notifications every 2 hours. However, participants were told to complete tasks only when they had a 5-minute distraction-free time window. We also discard the first instance of each task for each participant, as they are unfamiliar with the app at this point.

Figure 1. Screenshot of the Karolinska Sleepiness Scale and caffeination survey.

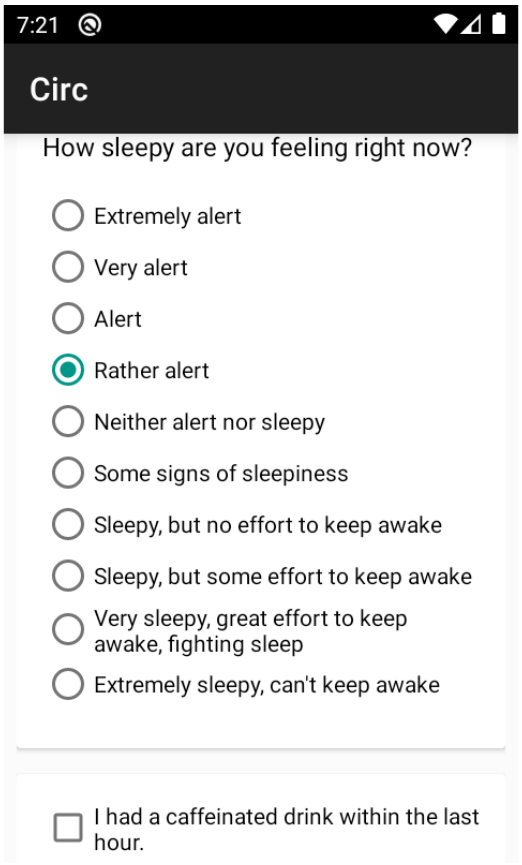


Figure 2. Screenshot of the PVT task.

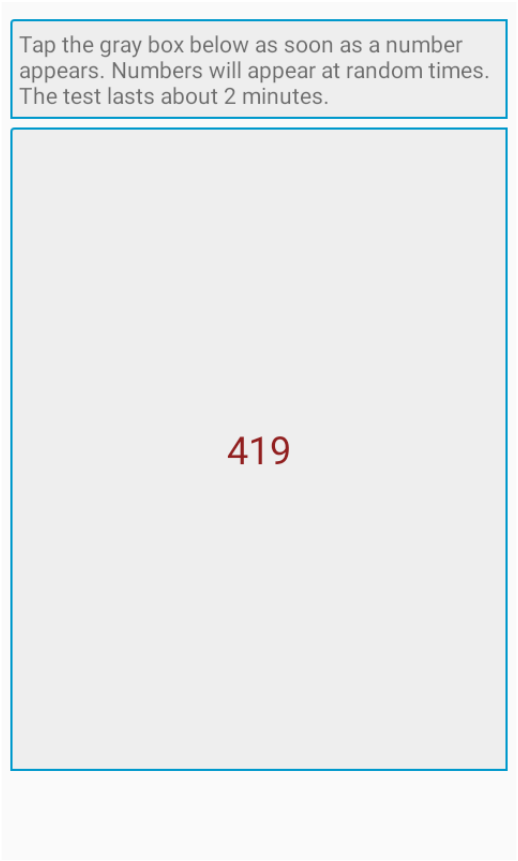
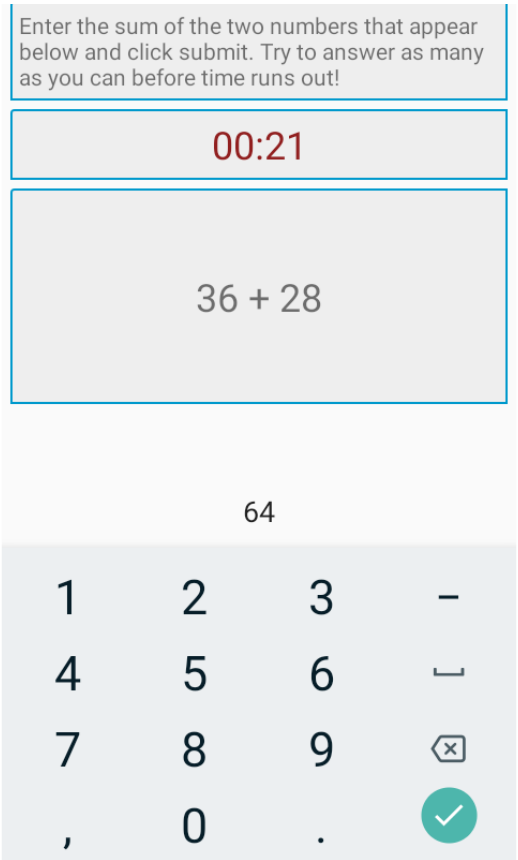


Figure 3. Screenshot of the ADD task.



Physiological Measures From Fitbit Fitness Tracker

We further collected physiological data from participants using a Fitbit Charge 3, which they were asked to wear on their nondominant hand throughout the study period. Participants were specifically instructed to wear the Fitbit while they slept and while they completed the PVT and ADD tests on their phones. They were also asked to either enable auto-sync via Bluetooth on their Fitbit app or to sync their trackers periodically so as to avoid loss of data.

We obtained participants' consent to collect data relating to their sleep, activity levels, and heart rate from Fitbit. Sleep data included the start and end time of each sleep session, time taken for sleep onset and waking up, and time spent in each stage of sleep (awake, light sleep, deep sleep, and rapid eye movement [REM] sleep) during each sleep session longer than 3 hours. Activity data included a minute-by-minute count of number of steps walked, distance covered, floors climbed, and calories burnt. Heart rate data included heart rate values each minute and the current day's resting heart rate.

Compliance and Compensation

For our analysis, we only used data from participants who had completed at least forty-two PVT and ADD tasks each over the entire study period (an average of at least one task per day). We excluded 8 participants through this criterion and were left with data from 1596 PVT and ADD sessions each. We describe our data set in more detail in the "Results" section.

Participants were compensated for granting access to their Fitbit data at the rate of US $10 per week, and for completing the PVT and ADD tasks on their smartphones at US $25 per week if they completed at least four sets of tasks each day. Monetary compensation was pro-rated for the period they contributed data if it was shorter than the study duration. Participants associated with the local police and fire departments were not offered monetary compensation in accordance with the departments' regulations. All participants were allowed to keep the Fitbit after the completion of the study. The study was approved by the Institutional Review Board of the University of Massachusetts Amherst.

Quantifying Chronobiological Sleep Metrics

Sleep Data and Components Considered

We use the sleep data obtained from Fitbit to calculate 2 sets of metrics corresponding to the circadian component *C* and the homeostatic component *H*. These components, along with the sleep inertia component *W*, collectively modulate cognitive performance in humans [2]. For simplicity, we ignore *W* in our analysis.

Circadian Component

The phase of the circadian component is regulated by the individual's chronotype [3]. To determine chronotype from Fitbit data, we leverage prior work on estimating individual chronotypes quantitatively using the mid-sleep point on free days (MSF), that is, the midpoint between sleep onset and wake up times on nonworking days [41]. As people tend to compensate for the sleep debt accumulated over work days by sleeping longer on free days, the quantitative measurement is typically adjusted accordingly. MSF is calculated as follows [24,41]:

$$MSF = MSF_{uc} - 0.5(SD_f - [N_w \times SD_w + N_f \times SD_f]/[N_w + N_f])$$

Here, MSF_{uc} is the uncorrected average mid-sleep point of the participant observed across the study duration. SD_f and SD_w represent the sleep duration on free and work days, while N_f and N_w are the number of free and work days, respectively. Based on this reference marker of the individual's circadian rhythm, an individual's internal time (InT) is defined as the time since the individual's MSF, or, in terms of the external (or "wall-clock") time ExT as InT=ExT–MSF. We also quantify the misalignment between the actual mid-sleep time of an individual and their MSF, hereafter referred to as sleep shift.

Homeostatic Component

The sleep homeostat is responsible for building up sleep pressure during wakefulness in a sigmoidal manner and then releasing this sleep pressure during recovery sleep sessions [2,9]. One metric that captures this sleep pressure is the time since waking up from the previous sleep session.

Further, to quantify whether enough recovery sleep has been obtained to release sleep pressure, we use 2 metrics, sleep need and sleep debt. We calculate sleep debt based on individual sleep need [4]. Sleep need SN for an individual is defined as

$$SN=\Sigma(SD_w \times N_w + SD_f \times N_f)/\Sigma(N_w + N_f)$$

where SD_w and SD_f are the sleep durations on workdays and free days, respectively; N_w is the number of work days; and N_f is the number of free days. Sleep debt accrued on a given night is defined as 1−(sleep duration/*SN*).

Results

Data Description

The data set collected as part of our study consisted of 923 nights of sleep data, 1,032,518 minutes of heart rate data over 813 days, and 1,169,510 minutes of activity data (calories burnt, steps walked, distance traveled, and floors climbed) over 903 days. In addition to these physiological data from the Fitbit, the 24 participants in our study also completed 2059 PVT and ADD tasks through our smartphone app over the study period. In this section, we describe the collected data in further detail.

Smartphone-Based Neurobehavioral Task Data

As described earlier, participants were asked to complete at least four PVT and ADD tasks each day, with at least two hours between each pair of consecutive tasks. The study app reminded users to complete these neurobehavioral tasks through push notifications every 2 hours. The average daily compliance rate across all participants was found to be 51%. We first filtered our data set to exclude all participants who had completed less than 42 PVT and ADD tasks throughout the 6-week study period (<1 task per day on average). This resulted in a data set consisting of 16 users who completed a total of 1596 tasks. The average number of tasks completed per day by these participants was 2.81 (SD 0.62). For the rest of the paper, we base our

analysis on the data from these participants alone. Figure 4 shows the average number of tasks completed by these participants each week of the study. While compliance rates dropped as the study progressed, the longitudinal nature of the study ensured that all participants completed a minimum of 47 tasks or more (mean 99.75 [SD 33.06] tasks).

Figure 4. Total number of tasks attempted per participant each week of the study. The markers show the mean across all participants and the error bars indicate the standard error of the mean.

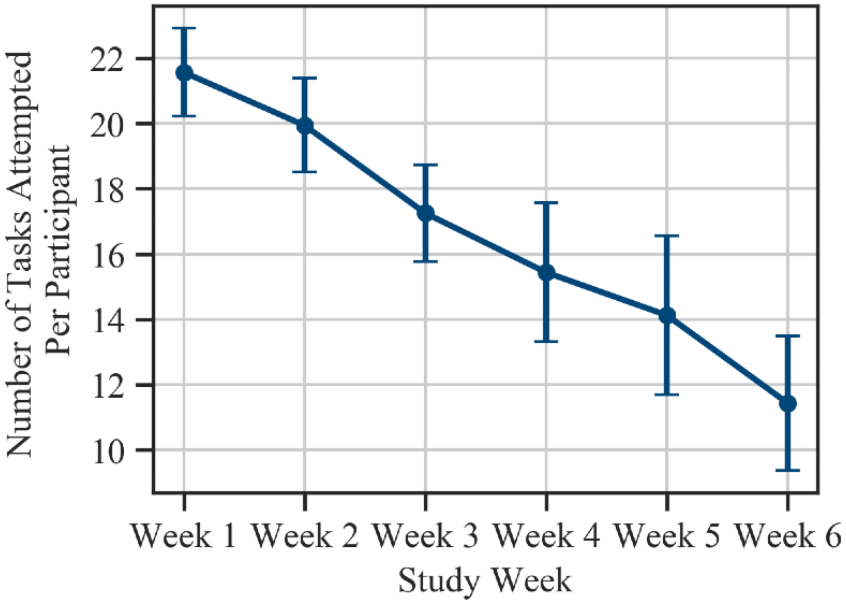


Table 1 shows the distribution of the task completion times in terms of time of day (or wall-clock time), internal time, and time since waking up. The participants' performance on these tasks is also reported—both PVT response times and ADD attempts were approximately normally distributed for each participant. Figure 5 shows the distribution of tasks completed by time of day.

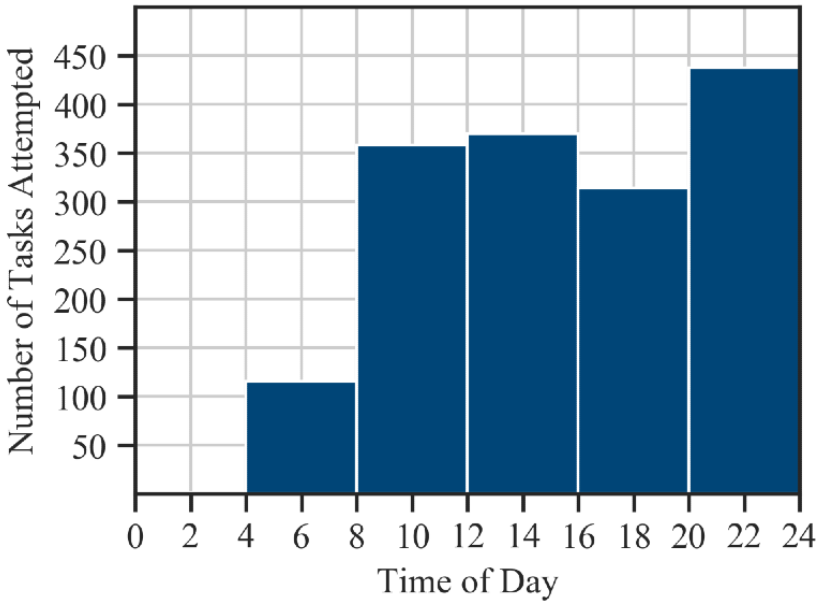
Table 1. Distribution of task completion timings and performance on PVT[a] and ADD[b] tasks.

Task completion timings and performance	N	Minimum	Maximum	Timing, mean (SD)
Time of task completion				
Time of day (hh:mm)	1596	3:22	23:58	15:34 (310.72 minutes)
Internal time (hh:mm)	1596	0:13	23:57	11:33 (318.19 minutes)
Time since waking up (minutes)	1596	1	1223	471.83 (309.75)
PVT task performance				
Median response time (milliseconds)	1596	234.0	798.0	354.53 (90.79)
Relative response time (%)	1596	–49.28	27.83	0.23 (8.93)
ADD task performance				
Number of additions attempted	1596	7.0	30.0	17.96 (3.95)
Relative number of additions attempted (%)	1596	–47.51	40.08	0.99 (13.46)

[a]PVT: Psychomotor Vigilance Test.

[b]ADD: Addition Test.

Figure 5. Distribution of PVT and ADD tasks attempted by time of day (binned into 2-hour intervals). Only one task was attempted in the 12AM-4AM interval.



Physiological Data From Fitbit

From the 923 sleep sessions collected, we first extracted sleep sessions preceding the 1596 neurobehavioral task instances described in the previous subsection. This resulted in a subset of 556 sleep sessions across 16 participants. The features extracted from these sleep sessions include sleep timings and duration, percentages of sleep sessions spent in different sleep stages (REM, light, and deep sleep), the duration of these sleep stages, and a Fitbit-provided sleep efficiency score. We use these metrics to further calculate chronobiological measures of sleep such as sleep debt and sleep shift. We also calculate average sleep measures over the last 7 days, imputing missing nights of sleep data with the average across all previous sleep sessions. The distribution of these sleep metrics is reported in Table 2. We investigate the effect of these sleep measures on median daytime alertness and cognitive throughput on the day following the sleep session.

Table 2. Distribution of sleep metrics obtained from the Fitbit as well as weekly averages of the same for all sleep sessions preceding the cognitive tasks.

Distribution of sleep metrics	N	Minimum	Maximum	Mean (SD)
Previous night's sleep metrics				
Sleep duration (minutes)	556	180.0	761.0	433.12 (87.47)
REM[a] sleep (%)	556	2.8	34.25	18.18 (5.32)
Light sleep (%)	556	30.04	73.71	52.92 (6.95)
Deep sleep (%)	556	0.0	33.57	15.92 (5.1)
Duration of REM sleep (minutes)	556	6.0	190.0	79.24 (29.04)
Duration of light sleep (minutes)	556	79.0	429.0	228.98 (54.13)
Duration of deep sleep (minutes)	556	0.0	157.0	68.69 (25.12)
Number of awake periods >5 minutes	556	0.0	8.0	3.03 (1.33)
Sleep efficiency (out of 100)	556	27	100	89.6 (14.11)
Sleep debt	556	–0.81	0.55	0.0 (0.18)
Sleep shift	556	–425.04	1199.21	7.1 (122.18)
Sleep metrics over last 1 week				
Average sleep duration (minutes)	556	221.0	564.25	433.27 (50.53)
Average REM sleep duration (minutes)	556	23.0	163.0	78.94 (15.59)
Average light sleep duration (minutes)	556	84.0	310.2	228.67 (36.2)
Average deep sleep duration (minutes)	556	28.0	123.5	69.14 (15.86)
Average sleep debt	556	–0.17	0.19	0.0 (0.06)
Average sleep shift	556	–311.47	202.93	6.97 (62.04)

[a]REM: rapid eye movement.

We also obtained minute-by-minute heart rate, calories burnt, steps walked, distance traveled, and floors climbed data from the participants' Fitbit trackers, along with an estimate of the current day's resting heart rate. Similar to sleep data, we extract heart rate and activity data corresponding to the time of completion of each of the PVT and ADD tasks. The distribution of these metrics at task time is represented in Table 3. We examined the effect of these momentary physiological measures and their aggregates over the last 60- and 10-minute intervals on both alertness and cognitive throughput.

Table 3. Distribution of heart rate and activity metrics obtained from the Fitbit during completion of the cognitive tasks.

Distribution of metrics	N	Minimum	Maximum	Mean (SD)
Heart rate metrics during PVT[a] and ADD[b] tasks				
Current heart rate (bpm)	1596	45.0	131.0	79.95 (13.81)
Resting heart rate (bpm)	1596	47.0	88.0	66.99 (9.14)
Activity metrics during PVT and ADD tasks				
Calories burnt	1596	0.77	9.12	1.48 (0.91)
Distance traveled (km)	1596	0.0	0.08	0.0 (0.01)
Number of steps walked	1596	0	112	2.35 (10.58)

[a]PVT: Psychomotor Vigilance Test.

[b]ADD: Addition Test.

Effect of Sleep on Daytime Alertness and Cognitive Throughput

In order to examine the effect of sleep on interday alertness and cognitive throughput, we first calculated the daily median response times and median number of additions attempted for each individual based on each day's completed tasks. We then calculate the relative daily response times and relative daily number of additions attempted for each day for each individual. These daily measures are calculated relative to the average scores of that individual across all days of participation (analogous to relative response times and relative numbers of additions attempted). We then examined the relationship between these relative cognitive measures and sleep features

from the Fitbit. We hypothesized that interday variations in alertness and cognitive throughput can be attributed to both the sleep timings and quality of the previous night's sleep session as well as sleep debt accumulated over a number of past sleep sessions.

To evaluate our hypothesis, we calculated the correlation between each sleep-related metric and the relative daily score. To reiterate, higher scores indicate higher alertness and cognitive throughput. In addition to sleep metrics from the previous night, we also considered cumulative sleep features over epochs of 1 week preceding the time the PVT/ADD test was administered. Because our data set consists of aggregated data from multiple participants, simple correlation can often produce spurious results due to violation of independence. To account for this as well as within- and inter-participant differences, we performed repeated measures correlation analysis [42] between our independent and dependent variables utilizing the Python package Pingouin [43]. This adjusts for interindividual variability using analysis of covariance, allowing us to draw population-level inferences while accounting for our repeated measures design.

We discuss the effect of each feature for which a significant (P<.05) repeated measures correlation (r_{rm}) was observed with alertness and cognitive throughput scores. We further juxtapose our findings from a noisy, real-world data set with existing theories on the effect of sleep on cognitive processes by discussing results of prior studies, most of which have been conducted in highly controlled laboratory settings.

Sleep Duration

When duration of sleep obtained during a recovery sleep session on a given night falls short of the individual's basal sleep need, it gives rise to the phenomenon of partial sleep deprivation (also known as sleep restriction or sleep loss) [44]. Sleep debt incurred in this manner can further exacerbate any existing long-term chronic sleep loss and the effects thereof [19].

Traditional sleep research has, however, focused far more on studying the effects of total sleep deprivation as compared to that of sleep restriction [44]. Total sleep deprivation occurs when individuals stay fully awake for long durations (typically 24-48 hours) with no recovery sleep obtained whatsoever. Mathematical models of cognitive performance have been heavily based on findings from such acute total sleep-deprivation studies (eg, [2,45]).

In in-the-wild studies such as ours, we are much more likely to observe chronic partial sleep deprivation, that is, a few hours of sleep loss each day, rather than acute total sleep deprivation. Chronic partial sleep deprivation is an increasingly common issue across populations due to increasing use of televisions, tablets, smartphones, laptops, or other electronic devices before bedtime [46-48]. Over prolonged exposure, the blue light from these screens suppresses the release of the sleep-inducing hormone melatonin, making it more difficult to fall asleep [49]. Therefore, it is imperative to characterize the effects of sleep debt incurred on a single, or multiple consecutive, night(s) on daytime performance measures.

Some recent studies have made efforts in this direction, studying the effects of chronic sleep loss on cognitive processes in controlled laboratory settings. For example, PVT performance has been shown to deteriorate with each consecutive day of sleep restriction [50-52]. Cognitive throughput, by contrast, has been demonstrated to show improvement across subsequent days which can be attributed to the effect of practice, implying that sleep restriction does not significantly impact cognitive throughput [50]. It has also been noted that while chronic sleep loss had a deteriorating effect on the ability to ignore distracting stimuli due to lower arousal levels, participants were able to overcome such effects on more cognitively complex logical reasoning tasks with additional effort [53].

The effect of recovery sleep to alleviate sleep debt–induced decline in performance has also been studied. PVT performance has been shown to improve with increasing hours of recovery sleep obtained, though there is some disagreement on whether this improvement is linear [54] or saturating exponential [55].

Table 4 shows the results from our study—we see that higher alertness sessions are observed following longer sleep duration over the previous night (repeated measures correlation r_{rm}=0.17, P<.001) and lower accumulated sleep debt over the previous week (r_{rm}=−0.1, P<.001). By contrast, we failed to observe any significant correlation between sleep duration–related metrics and cognitive throughput (P=.92 for sleep duration, P=.13 for previous week's sleep debt). This suggests that there is validity to the theory that individuals may be able to overcome any detrimental effects of sleep debt on more cognitively challenging tasks by potentially expending more effort (see [53]). Interestingly, this behavior has also been observed in the context of total sleep deprivation, where the "controlled attention model" [32] posits that tasks that are not intrinsically engaging or challenging are more affected by sleep loss. It is also worth investigating whether there continues to be no significant detrimental effect of cumulative sleep debt over a longer period on cognitive throughput.

Table 4. Effect of sleep duration on daytime alertness and cognitive throughput.

Feature	Effect on alertness		Effect on cognitive throughput	
	r_{rm}[a] (95% CI)	P value	r_{rm} (95% CI)	P value
Previous night's sleep duration	0.17 (0.08 to 0.25)	<.001	−0.004 (−0.09 to 0.08)	.92
Average nightly sleep debt incurred over previous week	−0.16 (−0.24 to −0.08)	<.001	0.06 (−0.02 to 0.15)	.13

[a]r_{rm}: repeated measures correlation coefficient.

Sleep Timing

Prior work has shown that sleeping in late, that is, a later sleep end time, can lead to higher sleep onset latency on the following night as well as higher daytime fatigue and sleepiness on subsequent days [56]. However, its effect on higher-order cognitive functions has not been studied previously. Even with a fixed wake-up time, a delayed bedtime leads to a sleep phase shift, which has been shown to produce a shift in salivary dim light melatonin onset—a marker of one's endogenous circadian rhythm [57]. We hypothesize that this circadian shift might adversely impact cognitive performance similar to a shift in wake-up times. As described earlier, sleep shift is calculated with respect to the participant's individual chronotype, or MSF.

Table 5 shows the findings from our study, which indicate that a phase shift in sleep sessions toward a later mid-sleep time corresponds to a decline in cognitive throughput the following day (r_{rm}=−0.13, P<.001). A later wake up time also has a similar effect on cognitive throughput (r_{rm}=−0.09, P=.03). However, alertness scores were not found to be significantly correlated with change in sleep timings (P=.09 for sleep shift, P=.74 for sleep end time).

Table 5. Effect of sleep timings on daytime alertness and cognitive throughput.

Feature	Effect on alertness		Effect on cognitive throughput	
	r_{rm}[a] (95% CI)	P value	r_{rm} (95% CI)	P value
Sleep shift	−0.07 (−0.16 to 0.01)	.09	−0.13 (−0.21 to −0.05)	<.001
Sleep end time	0.02 (−0.07 to 0.1)	.74	−0.09 (−0.18 to −0.01)	.03

[a]r_{rm}: repeated measures correlation coefficient.

Sleep Stages

There has been significant interest in understanding the effect of various sleep stages on daytime performance metrics, especially with the rising number of commercial wearables that claim to detect coarse-grained sleep stages. However, most experts advise caution in attributing performance to sleep stages inferred from these devices, and studies have shown that first-generation sleep trackers were generally quite poor at estimating sleep stages [58]. Nevertheless, prediction accuracies have shown improvement over time—for example, Fitbit Charge 2 showed 61% accuracy in detecting wake periods, 81% accuracy in detecting light sleep, 49% accuracy in detecting deep sleep, and 74% accuracy in detecting rapid eye movement (REM) sleep [59], whereas Garmin VivoFit 3 (Garmin International, Inc.) predicts deep, light, and REM sleep stages at roughly 69% accuracy rate, and predicts wake at 73% accuracy [60], which is a significant improvement over previous incarnations of these devices. As wearables get better at estimating sleep stages, it becomes more important to understand if we can leverage these insights to explain cognitive performance.

While the literature on the effect of sleep stages on cognitive performance is limited, studies have shown that electroencephalogram spindle density in non-REM sleep is a predictor of visual attention, verbal learning, and verbal fluency performance [61]. It has also been noted that light sleep, slow wave sleep, and REM sleep contribute to the recuperation of the dorsolateral prefrontal and inferior parietal cortices, which are areas involved in higher-order cognitive tasks [62].

Some studies have linked a reduction in total sleep duration specifically to a reduction in REM and stage 2 (light) sleep [63]. However, other studies have also claimed that selective REM sleep deprivation did not demonstrate changes in daytime sleepiness/alertness [64].

From Table 6, we see that there is a significant positive correlation between both REM (r_{rm}=0.12, P<.001) and light sleep (r_{rm}=0.15, P<.001) duration and daytime alertness. By contrast, sleep stages were not found to have a significant impact on cognitive throughput on the following day (P=.52 for REM sleep, P=.80 for light sleep). Duration of deep sleep was not found to have a significant impact on either alertness (P=.22) or cognitive throughput (P=.53).

Table 6. Effect of sleep stages on daytime alertness and cognitive throughput.

Feature	Effect on alertness		Effect on cognitive throughput	
	r_{rm}[a] (95% CI)	P value	r_{rm} (95% CI)	P value
Duration of rapid eye movement sleep	0.12 (0.04 to 0.2)	<.001	0.03 (−0.06 to 0.11)	.52
Duration of light sleep	0.15 (0.07 to 0.23)	<.001	0.01 (−0.07 to 0.09)	.80
Duration of deep sleep	0.05 (–0.03 to 0.14)	.22	−0.03 (−0.11 to 0.06)	.53

[a]r_{rm}: repeated measures correlation coefficient.

Factors Affecting Momentary Alertness and Cognitive Throughput

Having looked at the factors that influence day-to-day variations in daytime alertness and cognitive throughput, we now examine other factors that may impact momentary performance measures (ie, fluctuations in performance within a day). We specifically look into the effect of physical activity and heart rate, along with circadian and homeostatic effects.

Time of Day and Internal Time

Time of day has emerged as an important factor affecting multiple aspects of cognitive performance due to the circadian modulation of alertness and cognitive throughput [1]. Both alertness and cognitive throughput levels have been found to oscillate with a period approximately equal to 24 hours, with individuals achieving peak performance at similar times each day [65]. Further, it has been noted that an individual's chronotype influences the phase of this circadian modulation (also referred to as process *C*), thus determining the exact time at which this peak is observed [66].

Based on this well-established model of circadian fluctuations in performance, we would expect to see a roughly sinusoidal variation in PVT and ADD performance based on the time of day and participants' internal time. To examine whether such a relationship is indeed evident in our data, we fit cosinor models [67] with a period of 24 hours to relative response time and relative number of additions attempted based on both time of day and participants' internal time.

As seen in Table 7, both alertness and cognitive throughput are influenced by the current time, exhibiting an acrophase (circadian maxima) in the morning hours even if they do not coincide exactly. It is also important to note that interindividual differences, including chronotype, also influence these rhythms, which prompts us to study the variations in performance measures with respect to individuals' internal time. We see that the acrophase of these rhythms is reached about 13-14 hours after the individuals' MSF.

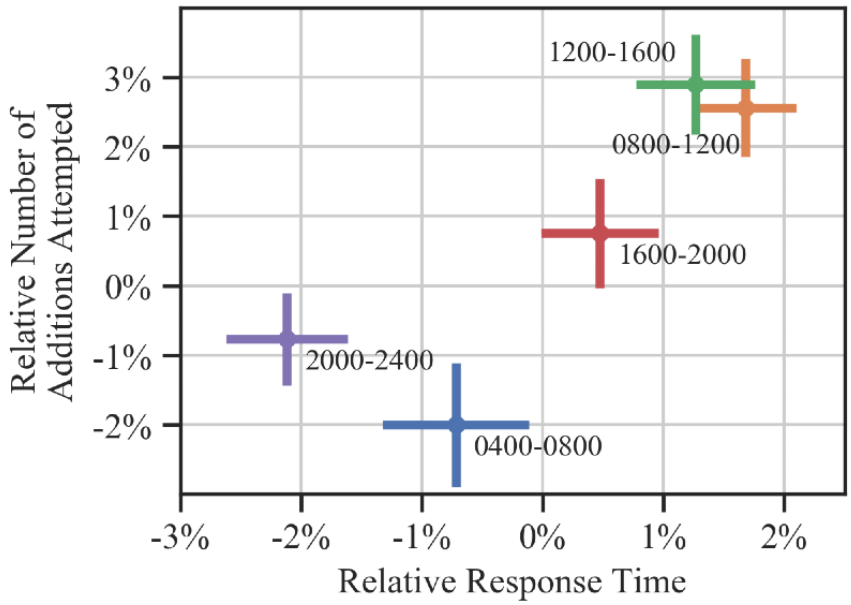
Table 7. Effect of time of day on momentary alertness and cognitive throughput.[a]

Feature	Effect on alertness	*P* value	Effect on cognitive throughput	*P* value
Time of day	Alertness varies sinusoidally with time of day, with an acrophase at 10:24 (09:26 to 11:21)	<.001	Cognitive throughput varies sinusoidally with time of day, with an acrophase at 09:19 (08:13 to 10:26)	<.001
Internal time	Alertness varies sinusoidally with internal time, with an acrophase at 14:24 (13:20 to 15:31)	<.001	Cognitive throughput varies sinusoidally with internal time, with an acrophase at 13:20 (12:02 to 14:38)	<.001

[a]The acrophase of the 24-hour circadian rhythm of cognitive performance is reported in hh:mm format with the corresponding 95% confidence intervals and *P* values.

Figure 6 visually illustrates the fluctuations in both performance measures across time of day. Here, we show the mean relative response times and number of additions attempted during each 4-hour time bin during the course of the day, omitting the 12:00-04:00 hours bin due to lack of data. We see that alertness increases through the day until around noon, after which it deteriorates through the evening. Cognitive throughput is lowest early in the morning, increases through the day peaking in the late afternoon, and decreases in the evening. In general, both alertness and cognitive throughput peak during the regular working hours of 08:00-16:00 hours.

Figure 6. Fluctuations in cognitive performance by time of day. The x-axis represents alertness in terms of relative response times, and the y-axis represents cognitive throughput in terms of relative number of additions attempted. The dots represent the mean scores in the marked time range while the error bars represent the standard error of the mean along each axis.



Homeostatic Pressure (ie, Time Since Waking Up)

The homeostatic component of sleep regulation is governed by a process *H* that induces sleep pressure (or "sleepiness") as a sigmoidal function of time since waking up from the last sleep session (typically previous night's sleep). During sleep, this pressure is released and *H* decays in a saturating exponential manner [9]. The homeostatic component of sleep regulation is known to affect alertness, and possibly cognitive throughput,

interacting with the circadian component described previously [2,15].

The effect of this sleep homeostat on daytime performance has been studied in terms of time awake in total sleep-deprivation studies, where participants are kept fully awake for very long periods (often up to 3 nights) [50]. The results of such studies generally indicate that alertness declines with an increase in hours of wakefulness [2,68,69], whereas there are conflicting theories about the effect of total sleep deprivation on more complex cognitive functions.

Traditionally, the effects of sleep deprivation were explained analogous to that of stress, based on the inverted-U model proposed by Yerkes and Dodson [70]. This "arousal" model essentially focused on the overall decline in arousal in sleep-deprived individuals in order to explain impairment of cognitive performance [71,72]. However, increasing empirical evidence showed that while performance on vigilance tasks decreased significantly across a night of sleep deprivation, performance did not vary significantly on more complex cognitive tasks [32,73]. This led to the proposal of several new theories to explain the effect of sleep deprivation on various cognitive functions.

Several researchers have proposed to single out vigilant attention as the cognitive process most susceptible to detrimental effects of sleep deprivation, while other cognitive tasks have varied sensitivity to [74], or a nonspecific effect of [75], sleep deprivation. Other studies have sought to explain their findings by claiming that tasks mediated by prefrontal cortex function are most impacted by homeostatic pressure [76,77]. Recent studies of neural activation patterns during cognitive tasks using functional magnetic resonance imaging have lent further support to this neuropsychological theory [78]. Another theory often used to explain sleep-deprivation effects is the controlled attention model [32], which posits that performance on tasks that require attentiveness and active engagement is less likely to be affected by sleep deprivation as compared to that on tasks that are not intrinsically interesting or engaging.

However, most of the aforementioned studies focus on the effects of sleep deprivation, that is, homeostatic pressure beyond at least one full day, which is much longer than typical homeostatic pressure in the working population. It is therefore still unclear how homeostatic pressure affects performance at points in time sooner after waking up, which is of greater relevance to us. To this end, we calculated the repeated measures correlation coefficient between the time since waking up from the last sleep session and performance on the corresponding PVT and ADD tasks.

As reported in Table 8, our analysis shows that alertness tends to decline with time since waking up (r_{rm}=−0.1, P<.001) while no significant effect is observed on cognitive throughput (P=.87). This finding agrees with observations from the early stages of sleep deprivation or constant routine protocol studies (eg, [65]). This suggests that cognitive throughput may indeed be affected less by sleep homeostasis (or the effect can be overcome by effort) or may be considerably affected by sleep inertia (ie, the drowsiness felt right after waking up).

Table 8. Effect of homeostatic pressure on momentary alertness and cognitive throughput.

Feature	Effect on alertness		Effect on cognitive throughput	
	r_{rm}[a] (95% CI)	*P* value	r_{rm} (95% CI)	*P* value
Time since waking up	−0.1 (−0.15 to −0.05)	<.001	0.004 (−0.05 to 0.05)	.87

[a]r_{rm}: repeated measures correlation coefficient.

Heart Rate

While there is limited prior research on understanding the impact of heart rate on cognitive performance, heart rhythm is generally thought to affect performance. In particular, there has been substantial work on understanding the relationship between HRV and alertness/cognitive performance [79]. HRV is considered a useful measure because it captures some aspects of the interplay between the sympathetic and parasympathetic nervous systems [80,81], which in turn has associations with the prefrontal cortex and hence cognitive performance [82].

Sleepiness is also known to relate to HRV—for example, Chua et al [83] showed that the R–R-interval power density correlates strongly with lapses on the PVT and can be used to estimate decrements in PVT performance caused by sleepiness. Henelius et al [12] report that HRV spectral power reflects vigilant attention in participants exposed to partial chronic sleep restriction. Heart rate measures have also been extensively studied in the context of fatigued driving, wherein several heart rate measures were found to be strong indicators of drowsiness under conditions of low mental workload [84]. Other work has reported that individuals with high HRV performed better on executive tasks compared to those with low HRV, but the 2 groups did not differ with regard to simple reaction time [11].

Recent work in ubiquitous computing research has also explored heart rhythm (and perceived heart rhythm) based interventions to improve cognitive performance—for example, Costa et al [85] showed that even changes in the perception of heart rate can lead to cognitive function improvement in an individual.

While our data set does not contain raw HRV information (this was not exposed by the device used for the study), we looked into the effects of coarser timescale variations in heart rate as well as direct heart rate measures on alertness and cognitive throughput. Table 9 reports the repeated measures correlation coefficients (along with the corresponding 95% confidence intervals) between various heart rate–based measures and both alertness and cognitive performance. We find that the current heart rate (measured minute-to-minute by the Fitbit) and the current day's resting heart rate are correlated with ADD performance (r_{rm}=0.16, P<.001), but a similar correlation with PVT scores was not observed (P=.05). Average heart rate values

over epochs of 10 and 60 minutes preceding the tasks were also found to be positively correlated with task performance, with the correlations being slightly higher with ADD scores than PVT scores (P<.001). Higher variance in heart rate (not to be confused with HRV) over the previous hour is significantly correlated with higher cognitive throughput (P=.03), while higher variance in a shorter interval of 10 minutes before task time corresponds to higher alertness measures (P=.04). While the correlations between variance in heart rate and task performance are low, they do further underscore the importance of investigating the effects of HRV as demonstrated by prior studies.

Table 9. Effect of heart rate on momentary alertness and cognitive throughput.

Feature	Effect on alertness		Effect on cognitive throughput	
	r_{rm} [a]	P value	r_{rm}	P value
Current heart rate	0.05 (0.0 to 0.1)	.05	0.16 (0.11 to 0.21)	<.001
Resting heart rate	−0.04 (−0.09 to 0.01)	.10	0.1 (0.05 to 0.15)	<.001
Average heart rate over last 60 minutes	0.11 (0.06 to 0.16)	<.001	0.16 (0.11 to 0.2)	<.001
Average heart rate over last 10 minutes	0.09 (0.04 to 0.14)	<.001	0.16 (0.11 to 0.21)	<.001
Variance in heart rate over last 60 minutes	0.04 (0.0 to 0.09)	.08	0.05 (0.0 to 0.1)	.03
Variance in heart rate over last 10 minutes	0.05 (0.0 to 0.1)	.04	0.05 (0.0 to 0.1)	.06

[a]r_{rm}: repeated measures correlation coefficient.

Physical Activity

The effect of physical activity on alertness and cognitive throughput has not been explored much in prior studies. Most in-laboratory studies prevent participants from engaging in any strenuous activity, while some sleep-deprivation studies use exercise as an additional stressor [14,86,87]. Within the latter category, studies remain inconsistent about the effects of exercise—Englund et al [14] claim that exercise did not compound effects of sleep loss, and physical activity may indeed delay any sleep loss–induced performance impairment on certain tasks. Angus et al [86] found that exercise neither increased nor decreased impairment caused by sleep deprivation. Exercise has also been found to decrease reaction times, but much less than a period of rest did [87].

The effect of physical activity on cognitive functioning of non-sleep-deprived individuals has been even less explored. Nevertheless, moderate-intensity exercise has been found to improve performance on information-processing tasks associated with sports [88,89]. These observations support the hypothesis that physically induced arousal due to exercise results in a performance improvement on cognitive tasks, which may not necessarily be explained by models of emotional arousal described previously.

Other studies suggest that physical activity can sometimes lead to higher alertness periods within break periods, but continued exertion causes this effect to wear off and may have an overall detrimental impact [13]. However, habitually engaging in moderate exercise and maintaining general fitness have also been correlated with better cognitive performance and academic achievement in children [90,91].

To investigate these effects in an in-the-wild setting, we examined whether current physical activity, or that undertaken in the recent past, has a momentary influence on either alertness or cognitive throughput. As shown in Table 10, we discovered a significant positive correlation of physical activity metrics aggregated over the previous 10- and 60-minute intervals with both PVT and ADD performance. This presents further empirical evidence in support of the arousal theory described previously.

Table 10. Effect of physical activity on momentary alertness and cognitive throughput.

Feature	Effect on alertness		Effect on cognitive throughput	
	r_{rm} [a]	*P* value	r_{rm}	*P* value
Average calories burnt over last 60 minutes	0.11 (0.06 to 0.16)	<.001	0.1 (0.05 to 0.15)	<.001
Average calories burnt over last 10 minutes	0.08 (0.03 to 0.12)	<.001	0.1 (0.05 to 0.15)	<.001
Variance in calories burnt over last 60 minutes	0.08 (0.03 to 0.13)	<.001	0.07 (0.02 to 0.12)	<.001
Variance in calories burnt over last 10 minutes	0.08 (0.03 to 0.13)	<.001	0.07 (0.02 to 0.12)	.01
Average distance traveled over last 60 minutes	0.08 (0.03 to 0.12)	<.001	0.07 (0.02 to 0.12)	<.001
Average distance traveled over last 10 minutes	0.06 (0.01 to 0.11)	.01	0.07 (0.02 to 0.12)	.01
Variance in distance traveled over last 60 minutes	0.07 (0.02 to 0.12)	.01	0.07 (0.02 to 0.12)	<.001
Variance in distance traveled over last 10 minutes	0.07 (0.03 to 0.12)	<.001	0.06 (0.01 to 0.11)	.01
Average steps walked over last 60 minutes	0.07 (0.02 to 0.12)	<.001	0.07 (0.02 to 0.12)	.01
Average steps walked over last 10 minutes	0.06 (0.01 to 0.11)	.02	0.07 (0.02 to 0.12)	<.001
Variance in steps walked over last 60 minutes	0.07 (0.02 to 0.12)	.01	0.06 (0.01 to 0.11)	.01
Variance in steps walked over last 10 minutes	0.08 (0.03 to 0.12)	<.001	0.07 (0.02 to 0.12)	.01

[a] r_{rm}: repeated measures correlation coefficient.

Discussion

Principal Findings

Our results indicate that performance on less engaging tasks that require sustained attention and that on tasks which are inherently more challenging are affected differently by various components of change in sleep patterns. Alertness is significantly influenced by the duration of sleep (*P*<.001) as well as time spent in various stages of sleep (*P*<.001 for both REM and light sleep), whereas cognitive throughput is moderated by phase shifts in sleep relative to an individual's internal circadian rhythm (*P*<.001). We also find that higher heart rate and physical activity preceding cognitively demanding tasks are positively correlated with better performance on said tasks.

Implications

Performance as a Multidimensional Metric

Our findings make a strong case for treating performance as a multidimensional metric and evaluating individuals' performance on multiple axes independently. In day-to-day societal functioning, different roles require different combinations of cognitive abilities in order to be responsibly and efficiently undertaken. For example, driving has been associated with a high need for vigilant alertness [52], whereas medical and emergency rescue personnel may rely more heavily on higher-order cognitive functions [92]. It is evident from our analysis that alertness and cognitive throughput are affected differently by different sleep-related variables, even though they covary similarly with respect to heart rate and physical activity. Thus, the ability to perform different tasks may be hindered differently due to the same changes in sleep patterns—a phenomenon that should be taken into account while predicting workplace performance, switching control of processes between human operator and automated assistants (in self-driving cars, for example), shift worker duty scheduling, etc.

Linking Self-Tracking Data to Actionable Insights

With commercial wearables claiming to accurately track an increasing number of physiological variables, there has been a growing interest in exploring the utility of these measures in drawing useful conclusions about users' physical, mental, and cognitive states. Researchers are continually seeking to answer the question of what can be learned from self-tracking data, and how this knowledge can be leveraged to close the loop by providing actionable feedback to the user [93].

Consumers of commercial wearable fitness trackers are seldom aware of distinctions in performance measures such as those described in this work, and manufacturers of wearable devices make frequent claims about how users can incorporate positive behavioral changes in their lifestyles based on metrics reported by these devices. For instance, Fitbit itself reports on its blog that exercise can boost happiness and engagement, thereby increasing productivity [23]. It also outlines steps to track an afternoon drop in attentiveness through the Fitbit app using sleep and activity data, encouraging users to then administer self-timed interventions such as caffeine [94]. The company also advises tracking and improving sleep to improve workplace productivity [22], though productivity is very loosely defined. The Fitbit app also provides personalized sleep insights, such as, "You sleep a bit better on nights after a run. It's subtle, but you spend 5 fewer minutes being restless/awake on those nights" [95]. However, the implications of these correlations are unclear—how does 5 fewer minutes of being restless/awake impact your day?

Our work shows that more comprehensive in-the-wild studies are required in order to meaningfully answer such questions. These studies should focus on understanding the effect of physiological data on specific, well-defined measures of daytime productivity and well-being, so that users can receive targeted

feedback on optimizing the individual measures that are personally of most consequence to them.

Limitations

Because our objective involved studying the effect of sleep-based metrics on cognitive performance, we opted for a longitudinal study wherein we could collect extensive data from a relatively small participant pool. A longitudinal study allowed us to capture a range of intra-individual variations and to get a better sense of our participants' natural sleep and performance rhythms over several weeks and ensure that the chronobiological sleep metrics such as mid-sleep time on free days are more stable. However, building predictive models of cognitive performance would require further study on larger populations in order to ensure wider generalizability. As a first step, we compared the distribution of alertness and cognitive throughput across our 3 subpopulations—shift workers, regular workers, and graduate students—and found no significant differences. Thus, we are optimistic that our observations would be replicable on larger and more diverse populations.

Our results are also fundamentally limited by the accuracy of the Fitbit fitness tracker in capturing physiological metrics. Nevertheless, recent studies on these devices are promising, showing significant improvement in the validity of their inferences over time [59]. In addition, while real-world physiological data may be noisy and is likely to be confounded by several external factors, our work demonstrates that there is substantial information in these signals that can be utilized toward modeling cognitive performance.

Conclusion

In conclusion, our work examines how metrics of sleep, activity, and heart rate that can be obtained from a commercial fitness tracker correlate with different facets of performance such as alertness and cognitive throughput. We achieve this through 2 complementary means—first, we delve into existing research in order to discern theories postulated on the basis of several controlled experiments. Second, we present insights from our own longitudinal in-the-wild study in an attempt to bridge the gap between laboratory findings and the real world.

We show that while alertness is sensitive to sleep duration as well as sleep stages, cognitive throughput exhibits no significant deteriorating effect from lack of sleep (P=.92). By contrast, irregular sleep timings have a significant effect on cognitive throughput (P<.001), but not on alertness. Both dimensions of cognitive performance show similar circadian fluctuations, but alertness is found to be more sensitive to homeostatic pressure. We also find that physical activity and heart rate have comparable effects on both alertness and cognitive throughput. The insights from our work make a strong case for treating performance as a multidimensional metric and evaluating individuals' performance on multiple axes.

Acknowledgments

This work is in part supported by the National Science Foundation under grant SBE 1839999 and start-up grant support from the College of Information and Computer Sciences and the Institute for Applied Life Sciences at the University of Massachusetts Amherst.

Conflicts of Interest

None declared.

References

1. Schmidt C, Collette F, Cajochen C, Peigneux P. A time to think: circadian rhythms in human cognition. Cogn Neuropsychol 2007 Oct;24(7):755-789. [doi: 10.1080/02643290701754158] [Medline: 18066734]
2. Jewett ME, Kronauer RE. Interactive mathematical models of subjective alertness and cognitive throughput in humans. J Biol Rhythms 1999 Dec 29;14(6):588-597. [doi: 10.1177/074873099129000920] [Medline: 10643756]
3. Juda M, Vetter C, Roenneberg T. Chronotype modulates sleep duration, sleep quality, and social jet lag in shift-workers. J Biol Rhythms 2013 Apr 19;28(2):141-151. [doi: 10.1177/0748730412475042] [Medline: 23606613]
4. Vetter C, Juda M, Roenneberg T. The influence of internal time, time awake, and sleep duration on cognitive performance in shiftworkers. Chronobiol Int 2012 Oct 13;29(8):1127-1138. [doi: 10.3109/07420528.2012.707999] [Medline: 22888791]
5. Martin JS, Hébert M, Ledoux E, Gaudreault M, Laberge L. Relationship of chronotype to sleep, light exposure, and work-related fatigue in student workers. Chronobiol Int 2012 Apr;29(3):295-304. [doi: 10.3109/07420528.2011.653656] [Medline: 22390242]
6. Medeiros ALD, Mendes DB, Lima PF, Araujo JF. The Relationships between Sleep-Wake Cycle and Academic Performance in Medical Students. Biological Rhythm Research 2010 Aug 09;32(2):263-270. [doi: 10.1076/brhm.32.2.263.1359]
7. Vetter C, Fischer D, Matera J, Roenneberg T. Aligning work and circadian time in shift workers improves sleep and reduces circadian disruption. Curr Biol 2015 Mar 30;25(7):907-911 [FREE Full text] [doi: 10.1016/j.cub.2015.01.064] [Medline: 25772446]
8. HIDALGO MPL. Performance of chronotypes on memory tests during the morning and the evening shifts. PR 2004 Aug;95(5):75. [doi: 10.2466/pr0.95.5.75-85]
9. Borbély AA. A two process model of sleep regulation. Hum Neurobiol 1982;1(3):195-204. [Medline: 7185792]

10. Akerstedt T, Folkard S. The three-process model of alertness and its extension to performance, sleep latency, and sleep length. Chronobiol Int 1997 Mar 07;14(2):115-123. [doi: 10.3109/07420529709001149] [Medline: 9095372]
11. Hansen AL, Johnsen BH, Thayer JF. Vagal influence on working memory and attention. International Journal of Psychophysiology 2003 Jun;48(3):263-274. [doi: 10.1016/s0167-8760(03)00073-4]
12. Henelius A, Sallinen M, Huotilainen M, Müller K, Virkkala J, Puolamäki K. Heart rate variability for evaluating vigilant attention in partial chronic sleep restriction. Sleep 2014 Jul 01;37(7):1257-1267 [FREE Full text] [doi: 10.5665/sleep.3850] [Medline: 24987165]
13. O'Neill TR, Krueger GP, Van Hemel SB, McGowan AL, Rogers WC. Effects of Cargo Loading and Unloading on Truck Driver Alertness. Transportation Research Record 1999 Jan;1686(1):42-48. [doi: 10.3141/1686-07]
14. Englund CE, Ryman DH, Naitoh P, Hodgdon JA. Cognitive performance during successive sustained physical work episodes. Behavior Research Methods, Instruments, & Computers 1985 Jan;17(1):75-85. [doi: 10.3758/bf03200899]
15. Basner M, Rao H, Goel N, Dinges DF. Sleep deprivation and neurobehavioral dynamics. Current Opinion in Neurobiology 2013 Oct;23(5):854-863. [doi: 10.1016/j.conb.2013.02.008]
16. Cajochen C, Münch M, Kobialka S, Kräuchi K, Steiner R, Oelhafen P, et al. High sensitivity of human melatonin, alertness, thermoregulation, and heart rate to short wavelength light. J Clin Endocrinol Metab 2005 Mar;90(3):1311-1316. [doi: 10.1210/jc.2004-0957] [Medline: 15585546]
17. Wesensten NJ, Killgore WDS, Balkin TJ. Performance and alertness effects of caffeine, dextroamphetamine, and modafinil during sleep deprivation. J Sleep Res 2005 Sep;14(3):255-266 [FREE Full text] [doi: 10.1111/j.1365-2869.2005.00468.x] [Medline: 16120100]
18. Philip P, Sagaspe P, Taillard J, Moore N, Guilleminault C, Sanchez-Ortuno M, et al. Fatigue, sleep restriction, and performance in automobile drivers: a controlled study in a natural environment. Sleep 2003 May 01;26(3):277-280. [doi: 10.1093/sleep/26.3.277] [Medline: 12749545]
19. Van Dongen HP, Rogers NL, Dinges DF. Sleep debt: Theoretical and empirical issues. Sleep and Biological Rhythms 2003 Feb;1(1):5-13. [doi: 10.1046/j.1446-9235.2003.00006.x]
20. Johnson M, Duffy J, Dijk D, Ronda J, Dyal C, Czeisler C. Short-term memory, alertness and performance: a reappraisal of their relationship to body temperature. J Sleep Res 1992 Mar;1(1):24-29 [FREE Full text] [doi: 10.1111/j.1365-2869.1992.tb00004.x] [Medline: 10607021]
21. Moonshot Factory - 2019 Stanford Medicine Big Data | Precision Health - YouTube Internet. Jonathan Berent, X. URL: https://www.youtube.com/watch?v=5HenODfjEHg [accessed 2020-08-26]
22. Fitbit Blog. Getting More Sleep Can Help You Be an All-Star at Work Internet. 2017. URL: https://blog.fitbit.com/sleep-work/ [accessed 2020-08-26]
23. Fitbit Blog. 5 Easy Ways to Be More Productive at Work Internet. 2017. URL: https://blog.fitbit.com/5-easy-ways-productive-work/ [accessed 2020-08-26]
24. Abdullah S, Murnane E, Matthews M, Kay M, Kientz J, Gay G, et al. Cognitive rhythms: unobtrusive and continuous sensing of alertness using a mobile phone. : ACM; 2016 Presented at: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 12-16, 2016; Heidelberg Germany.
25. Dinges DF, Powell JW. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behavior Research Methods, Instruments, & Computers 1985 Nov;17(6):652-655. [doi: 10.3758/bf03200977]
26. Mark G, Iqbal S, Czerwinski M, Johns P. Bored mondays and focused afternoons: the rhythm of attention and online activity in the workplace. New York, NY: ACM Press; 2014 Presented at: Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI 14; April 26-May 1, 2014; Toronto, ON, Canada.
27. Mark G, Wang YOI, Niiya M, Reich S. Sleep Debt in Student Life: Online Attention Focus, Facebook, and Mood. New York, NY: ACM Press; 2016 Presented at: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; 2016; San Jose, CA.
28. Wahl F, Amft O. Data and Expert Models for Sleep Timing and Chronotype Estimation from Smartphone Context Data and Simulations. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2018 Sep 18;2(3):1-28.
29. Althoff T, Horvitz E, White R, Zeitzer J. Harnessing the Web for Population-Scale Physiological Sensing: A Case Study of Sleep and Performance. Geneva, Switzerland: International World Wide Web Conferences Steering Committee; 2017 Presented at: Proceedings of the 26th International Conference on World Wide Web; 2017; Perth, WA, Australia.
30. Murnane E, Abdullah S, Matthews M, Choudhury T, Gay G. Social (media) jet lag: how usage of social technology can modulate and reflect circadian rhythms. New York, NY: ACM Press; 2015 Presented at: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp 15; 2015; Osaka, Japan.
31. Murnane E, Abdullah S, Matthews M, Kay M, Kientz J, Choudhury T, et al. Mobile manifestations of alertness: connecting biological rhythms with patterns of smartphone app use. New York, NY: ACM Press; 2016 Presented at: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services; 2016; Florence, Italy.
32. Pilcher J, Band D, Odle-Dusseau H, Muth E. Human performance under sustained operations and acute sleep deprivation conditions: toward a model of controlled attention. Aviat Space Environ Med 2007 May;78(5 Suppl):B15-B24. [Medline: 17547301]

33. Dingler T, Schmidt A, Machulla T. Building Cognition-Aware Systems. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-15 [FREE Full text] [doi: 10.1145/3132025]
34. Åkerstedt T, Gillberg M. Subjective and Objective Sleepiness in the Active Individual. International Journal of Neuroscience 2009 Jul 07;52(1-2):29-37. [doi: 10.3109/00207459008994241]
35. Dijk D, Duffy J, Czeisler C. Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. Journal of Sleep Research 1992 Jun;1(2):112-117. [doi: 10.1111/j.1365-2869.1992.tb00021.x]
36. Klein K, Wegmann H, Athanassenas G, Hohlweck H, Kuklinski P. Air operations and circadian performance rhythms. Aviat Space Environ Med 1976 Mar;47(3):221-230. [Medline: 769776]
37. Grant DA, Honn KA, Layton ME, Riedy SM, Van Dongen HPA. 3-minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation. Behav Res Methods 2017 Jun 20;49(3):1020-1029. [doi: 10.3758/s13428-016-0763-8] [Medline: 27325169]
38. Honn KA, Riedy SM, Grant DA. Validation of a Portable, Touch-Screen Psychomotor Vigilance Test. Aerospace Medicine and Human Performance 2015 May 01;86(5):428-434. [doi: 10.3357/amhp.4165.2015]
39. Kay M, Rector K, Consolvo S, Greenstein B, Wobbrock J, Watson N, et al. PVT-Touch: Adapting a Reaction Time Test for Touchscreen Devices. New York, NY: IEEE; 2013 Presented at: Proceedings of the ICTs for improving Patients Rehabilitation Research Techniques; 2013; Venice, Italy.
40. Basner M, Dinges D. Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. Sleep 2011 May 01;34(5):581-591 [FREE Full text] [doi: 10.1093/sleep/34.5.581] [Medline: 21532951]
41. Roenneberg T, Wirz-Justice A, Merrow M. Life between clocks: daily temporal patterns of human chronotypes. J Biol Rhythms 2003 Feb;18(1):80-90. [doi: 10.1177/0748730402239679] [Medline: 12568247]
42. Bakdash JZ, Marusich LR. Repeated Measures Correlation. Front. Psychol 2017 Apr 7;8:456. [doi: 10.3389/fpsyg.2017.00456]
43. Vallat R. Pingouin: statistics in Python. JOSS 2018 Nov;3(31):1026. [doi: 10.21105/joss.01026]
44. Reynolds A, Banks S. Total sleep deprivation, chronic sleep restriction and sleep disruption. Prog Brain Res 2010;185:91-103. [doi: 10.1016/B978-0-444-53702-7.00006-3] [Medline: 21075235]
45. Akerstedt T, Folkard S. Validation of the S and C components of the three-process model of alertness regulation. Sleep 1995 Jan;18(1):1-6. [doi: 10.1093/sleep/18.1.1] [Medline: 7761737]
46. Cajochen C, Frey S, Anders D, Späti J, Bues M, Pross A, et al. Evening exposure to a light-emitting diodes (LED)-backlit computer screen affects circadian physiology and cognitive performance. J Appl Physiol (1985) 2011 May;110(5):1432-1438 [FREE Full text] [doi: 10.1152/japplphysiol.00165.2011] [Medline: 21415172]
47. Shochat T. Impact of lifestyle and technology developments on sleep. NSS 2012 Mar:19. [doi: 10.2147/nss.s18891]
48. Twenge JM, Krizan Z, Hisler G. Decreases in self-reported sleep duration among U.S. adolescents 2009-2015 and association with new media screen time. Sleep Med 2017 Nov;39:47-53. [doi: 10.1016/j.sleep.2017.08.013] [Medline: 29157587]
49. Tosini G, Ferguson I, Tsubota K. Effects of blue light on the circadian system and eye physiology. Mol Vis 2016;22:61-72 [FREE Full text] [Medline: 26900325]
50. Van Dongen HPA, Maislin G, Mullington J, Dinges D. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 2003 Mar 15;26(2):117-126. [doi: 10.1093/sleep/26.2.117] [Medline: 12683469]
51. Dinges D, Pack F, Williams K, Gillen K, Powell J, Ott G, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep 1997 Apr;20(4):267-277. [Medline: 9231952]
52. Jackson M, Croft R, Kennedy G, Owens K, Howard M. Cognitive components of simulated driving performance: Sleep loss effects and predictors. Accid Anal Prev 2013 Jan;50:438-444. [doi: 10.1016/j.aap.2012.05.020] [Medline: 22721550]
53. Blagrove M, Alexander C, Horne JA. The effects of chronic sleep reduction on the performance of cognitive tasks sensitive to sleep deprivation. Appl. Cognit. Psychol 1995 Feb;9(1):21-40. [doi: 10.1002/acp.2350090103]
54. Banks S, Van DH, Maislin G, Dinges D. Neurobehavioral Dynamics Following Chronic Sleep Restriction: Dose-Response Effects of One Night for Recovery. Sleep 2010 Aug;? 2010;33(8):1026. [doi: 10.1093/sleep/33.8.1013]
55. Jewett M, Dijk D, Kronauer R, Dinges D. Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. Sleep 1999 Mar 15;22(2):171-179. [doi: 10.1093/sleep/22.2.171] [Medline: 10201061]
56. Taylor A, Wright HR, Lack LC. Sleeping-in on the weekend delays circadian phase and increases sleepiness the following week. Sleep Biol Rhythms 2008;6(3):172-179. [doi: 10.1111/j.1479-8425.2008.00356.x]
57. Burgess HJ, Eastman CI. Early versus late bedtimes phase shift the human dim light melatonin rhythm despite a fixed morning lights on time. Neurosci Lett 2004 Feb 12;356(2):115-118 [FREE Full text] [doi: 10.1016/j.neulet.2003.11.032] [Medline: 14746877]
58. Mantua J, Gravel N, Spencer RMC. Reliability of Sleep Measures from Four Personal Health Monitoring Devices Compared to Research-Based Actigraphy and Polysomnography. Sensors (Basel) 2016 May 05;16(5):646 [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]
59. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. Chronobiol Int 2018 Apr;35(4):465-476. [doi: 10.1080/07420528.2017.1413578] [Medline: 29235907]

60. Garmin Blog. Garmin Health Announces Sleep Study Results. 2019. URL: https://www.garmin.com/en-US/blog/health/garmin-health-announces-sleep-study-results/ [accessed 2020-08-26]
61. Lafortune M, Gagnon J, Martin N, Latreille V, Dubé J, Bouchard M, et al. Sleep spindles and rapid eye movement sleep as predictors of next morning cognitive performance in healthy middle-aged and older participants. J Sleep Res 2014 Apr 18;23(2):159-167 [FREE Full text] [doi: 10.1111/jsr.12108] [Medline: 24245769]
62. Thomas M, Sing H, Belenky G, Holcomb H, Mayberg H, Dannals R, et al. Neural basis of alertness and cognitive performance impairments during sleepiness. I. Effects of 24 h of sleep deprivation on waking human regional brain activity. J Sleep Res 2000 Dec 18;9(4):335-352 [FREE Full text] [doi: 10.1046/j.1365-2869.2000.00225.x] [Medline: 11123521]
63. Kecklund G, Akerstedt T, Lowden A. Morning work: effects of early rising on sleep and alertness. Sleep 1997 Mar;20(3):215-223. [doi: 10.1093/sleep/20.3.215] [Medline: 9178917]
64. Nykamp K, Rosenthal L, Folkerts M, Roehrs T, Guido P, Roth T. The effects of REM sleep deprivation on the level of sleepiness/alertness. Sleep 1998 Sep 15;21(6):609-614. [doi: 10.1093/sleep/21.6.609] [Medline: 9779520]
65. Cajochen C, Khalsa SBS, Wyatt JK, Czeisler CA, Dijk D. EEG and ocular correlates of circadian melatonin phase and human performance decrements during sleep loss. Am J Physiol 1999 Sep 01;277(3 Pt 2):R640-R649. [doi: 10.1152/ajpregu.1999.277.3.r640] [Medline: 10484479]
66. Horne JA, Brass CG, Pettitt AN. Ciradian performance differences between morning and evening "types". Ergonomics 1980 Jan 27;23(1):29-36. [doi: 10.1080/00140138008924715] [Medline: 7189149]
67. Cornelissen G. Cosinor-based rhythmometry. Theor Biol Med Model 2014 Apr 11;11:16 [FREE Full text] [doi: 10.1186/1742-4682-11-16] [Medline: 24725531]
68. Graw P, Kräuchi K, Knoblauch V, Wirz-Justice A, Cajochen C. Circadian and wake-dependent modulation of fastest and slowest reaction times during the psychomotor vigilance task. Physiol Behav 2004 Feb;80(5):695-701. [doi: 10.1016/j.physbeh.2003.12.004] [Medline: 14984804]
69. Lim J, Dinges D. Sleep deprivation and vigilant attention. Ann N Y Acad Sci 2008;1129(1):305-322. [doi: 10.1196/annals.1417.002] [Medline: 18591490]
70. Yerkes RM, Dodson JD. The relation of strength of stimulus to rapidity of habit-formation. J. Comp. Neurol. Psychol 1908 Nov;18(5):459-482. [doi: 10.1002/cne.920180503]
71. Wilkinson RT. Interaction of lack of sleep with knowledge of results, repeated testing, and individual differences. J Exp Psychol 1961 Sep;62(3):263-271. [doi: 10.1037/h0048787] [Medline: 14007063]
72. Williams HL, Lubin A, Goodnow JJ. Impaired performance with acute sleep loss. Psychological Monographs: General and Applied 1959;73(14):1-26. [doi: 10.1037/h0093749]
73. Quigley N, Green JF, Morgan D, Idzikowski C, King DJ. The effect of sleep deprivation on memory and psychomotor function in healthy volunteers. Hum Psychopharmacol 2000 May;15(3):171-177. [doi: 10.1002/(SICI)1099-1077(200004)15:3<171::AID-HUP155>3.0.CO;2-D] [Medline: 12404330]
74. Durmer JS, Dinges DF. Neurocognitive consequences of sleep deprivation. Semin Neurol 2005 Mar;25(1):117-129. [doi: 10.1055/s-2005-867080] [Medline: 15798944]
75. Balkin TJ, Rupp T, Picchioni D, Wesensten NJ. Sleep loss and sleepiness: current issues. Chest 2008 Oct;134(3):653-660. [doi: 10.1378/chest.08-1064] [Medline: 18779203]
76. Harrison Y, Horne J, Rothwell A. Prefrontal neuropsychological effects of sleep deprivation in young adults--a model for healthy aging? Sleep 2000 Dec 15;23(8):1067-1073. [Medline: 11145321]
77. Jones K, Harrison Y. Frontal lobe function, sleep loss and fragmented sleep. Sleep Med Rev 2001 Dec;5(6):463-475. [doi: 10.1053/smrv.2001.0203] [Medline: 12531154]
78. Chee MWL. Functional Imaging of Working Memory after 24 Hr of Total Sleep Deprivation. Journal of Neuroscience 2004 May 12;24(19):4560-4567. [doi: 10.1523/jneurosci.0007-04.2004]
79. O'Hanlon J. Heart Rate Variability: A New Index of Driver Alertness/Fatigue. SAE Technical Paper 720141 1972:1-8. [doi: 10.4271/720141]
80. Goldberger JJ, Challapalli S, Tung R, Parker MA, Kadish AH. Relationship of heart rate variability to parasympathetic effect. Circulation 2001 Apr 17;103(15):1977-1983. [doi: 10.1161/01.cir.103.15.1977] [Medline: 11306527]
81. Ernst G. Heart-Rate Variability-More than Heart Beats? Front Public Health 2017 Sep 11;5:240 [FREE Full text] [doi: 10.3389/fpubh.2017.00240] [Medline: 28955705]
82. Thayer JF, Hansen AL, Saus-Rose E, Johnsen BH. Heart rate variability, prefrontal neural function, and cognitive performance: the neurovisceral integration perspective on self-regulation, adaptation, and health. Ann Behav Med 2009 Apr 8;37(2):141-153. [doi: 10.1007/s12160-009-9101-z] [Medline: 19424767]
83. Chua E, Tan W, Yeo S, Lau P, Lee I, Mien I, et al. Heart rate variability can be used to estimate sleepiness-related decrements in psychomotor vigilance during total sleep deprivation. Sleep 2012 Mar 01;35(3):325-334 [FREE Full text] [doi: 10.5665/sleep.1688] [Medline: 22379238]
84. Magjarevic R, Heinze C, Trutschel U, Schnupp T, Sommer D, Schenka A, et al. Operator Fatigue Estimation Using Heart Rate Measures. Berlin, Heidelberg: Springer; 2009 Presented at: World Congress on Medical Physics and Biomedical Engineering; September 7-12, 2009; Munich, Germany.

85. Costa J, Guimbretière F, Jung MF, Choudhury T. BoostMeUp. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2019 Jun 21;3(2):1-23. [doi: 10.1145/3328911]
86. Angus RG, Heslegrave RJ, Myles WS. Effects of prolonged sleep deprivation, with and without chronic physical exercise, on mood and performance. Psychophysiology 1985 May;22(3):276-282. [doi: 10.1111/j.1469-8986.1985.tb01601.x] [Medline: 4011797]
87. LeDuc P, Caldwell J, Ruyak P. The effects of exercise versus napping on alertness and mood in sleep-deprived aviators. US Aeromedical Research Laboratory Technical Report. 2000. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.459.2378&rep=rep1&type=pdf [accessed 2021-02-14]
88. Davranche K, Audiffren M. Facilitating effects of exercise on information processing. J Sports Sci 2004 May;22(5):419-428. [doi: 10.1080/02640410410001675289] [Medline: 15160595]
89. McMorris T, Graydon J. The effect of exercise on cognitive performance in soccer-specific tests. J Sports Sci 1997 Oct;15(5):459-468. [doi: 10.1080/026404197367092] [Medline: 9386203]
90. Keeley TJ, Fox KR. The impact of physical activity and fitness on academic achievement and cognitive performance in children. International Review of Sport and Exercise Psychology 2009 Sep;2(2):198-214. [doi: 10.1080/17509840903233822]
91. Ploughman M. Exercise is brain food: the effects of physical activity on cognitive function. Dev Neurorehabil 2008 Jul 10;11(3):236-240. [doi: 10.1080/17518420801997007] [Medline: 18781504]
92. Dula DJ, Dula NL, Hamrick C, Wood G. The effect of working serial night shifts on the cognitive functioning of emergency physicians. Ann Emerg Med 2001 Aug;38(2):152-155. [doi: 10.1067/mem.2001.116024] [Medline: 11468610]
93. van Berkel N, Luo C, Ferreira D, Goncalves J, Kostakos V. The curse of quantified-self: an endless quest for answers. New York, NY: ACM Press; 2015 Presented at: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers - UbiComp 15; 2015; Osaka, Japan.
94. Fitbit Blog. Fitbit Data Can Help You Beat Afternoon Energy Crashes Internet. 2017. URL: https://blog.fitbit.com/afternoon-slump/ [accessed 2020-08-26]
95. Get a Good Night's Sleep with Fitbit's Sleep Tools. Fitbit Blog. 2019. URL: https://blog.fitbit.com/get-a-good-nights-sleep-with-fitbits-new-sleep-tools/ [accessed 2021-02-12]

Abbreviations

ADD: Addition Test
MSF: mid-sleep point on free days
PVT: Psychomotor Vigilance Test
REM: rapid eye movement

Edited by G Eysenbach, R Kukafka; submitted 28.08.20; peer-reviewed by T Tsoneva; comments to author 22.09.20; revised version received 28.09.20; accepted 20.01.21; published 18.02.21.

Please cite as:
Kalanadhabhatta M, Rahman T, Ganesan D
Effect of Sleep and Biobehavioral Patterns on Multidimensional Cognitive Performance: Longitudinal, In-the-Wild Study
J Med Internet Res 2021;23(2):e23936
URL: http://www.jmir.org/2021/2/e23936/
doi:10.2196/23936
PMID:33599622

Original Paper

The Epidemiology of Patients' Email Addresses in a French University Hospital: Case-Control Study

Vincent Looten[1,2,3], MD; Antoine Neuraz[2,3,4], MD, PhD; Nicolas Garcelon[2,5], PhD; Anita Burgun[1,2,3,4], MD, PhD; Gilles Chatellier[1], MD, PhD; Bastien Rance[1,2,3], PhD

[1]Medical Informatics Department, Hôpital Européen Georges-Pompidou, Assistance Publique-Hôpitaux de Paris, Paris, France

[2]UMRS 1138 - Centre de Recherche des Cordeliers, Université Paris Descartes, Sorbonne Paris Cité, INSERM, Paris, France

[3]Université Paris Descartes, Paris, France

[4]Department of Medical Informatics, Hôpital Necker-Enfant Malades, Assistance Publique des Hôpitaux de Paris, Paris, France

[5]Institut Imagine, Université Paris Descartes, Université Paris Descartes-Sorbonne Paris Cité, Paris, France

Corresponding Author:
Bastien Rance, PhD
Medical Informatics Department
Hôpital Européen Georges-Pompidou
Assistance Publique-Hôpitaux de Paris
20 rue Leblanc
Paris
France
Phone: 33 156095985
Email: bastien.rance@aphp.fr

Abstract

Background: Health care professionals are caught between the wish of patients to speed up health-related communication via emails and the need for protecting health information.

Objective: We aimed to analyze the demographic characteristics of patients providing an email, and study the distribution of emails' domain names.

Methods: We used the information system of the European Hospital Georges Pompidou (HEGP) to identify patients who provided an email address. We used a 1:1 matching strategy to study the demographic characteristics of the patients associated with the presence of an email, and described the characteristics of the emails used (in terms of types of emails—free, business, and personal).

Results: Overall, 4.22% (41,004/971,822) of the total population of patients provided an email address. The year of last contact with the patient is the strongest driver of the presence of an email address (odds ratio [OR] 20.8, 95% CI 18.9-22.9). Patients more likely to provide an email address were treated for chronic conditions and were more likely born between 1950 and 1969 (taking patients born before 1950 as reference [OR 1.60, 95% CI 1.54-1.67], and compared to those born after 1990 [OR 0.56, 95% CI 0.53-0.59]). Of the 41,004 email addresses collected, 37,779 were associated with known email providers, 31,005 email addresses were associated with Google, Microsoft, Orange, and Yahoo!, 2878 with business emails addresses, and 347 email addresses with personalized domain names.

Conclusions: Emails have been collected only recently in our institution. The importance of the year of last contact probably reflects this recent change in contact information collection policy. The demographic characteristics and especially the age distribution are likely the result of a population bias in the hospital: patients providing email are more likely to be treated for chronic diseases. A risk analysis of the use of email revealed several situations that could constitute a breach of privacy that is both likely and with major consequences. Patients treated for chronic diseases are more likely to provide an email address, and are also more at risk in case of privacy breach. Several common situations could expose their private information. We recommend a very restrictive use of the emails for health communication.

(J Med Internet Res 2021;23(2):e13992) doi:10.2196/13992

KEYWORDS

email; data privacy; health communication

Introduction

Background

In the digital era, there is a natural worldwide trend for patients and physicians to use modern communication tools including emails, texting, social networks, and instant messaging services. Several studies illustrate this trend throughout the world [1-4]. The arrival of *digital natives* [5] among both the health care workforce and the adult population, alongside the development of mobile apps and wearable sensors, is probably going to amplify that trend. However, the digital sharing of health information has strong implications in terms of data privacy, data governance, and data security. Focusing on emails only, several questions can be raised: (1) in terms of data privacy, both the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe implement rules regarding the use of emails to exchange medical information. Regardless of these regulations, the fact that patients use either personal or business email accounts raises specific questions in itself. (2) In terms of data governance, the use of cloud-based providers or foreign providers depending on foreign jurisdiction raises questions related to the sovereignty of citizens' data. (3) Finally, many data privacy breaches have shown how considerable the impact of security on the confidentiality of personal data is (eg, the recent Singapore health data breach [6], or the revelations on the misuse of Facebook data [7]).

Regarding data privacy and electronic communication, the American HIPAA Compliance Guide [8] clearly states that sending private medical messages by SMS text messages or email requires a secure platform following the HIPAA safeguards and therefore should not be done using unsecured usual emails. Similarly, in Europe, health care professionals are not supposed to send health-related information using nonsecured platforms [9].

Regardless of those guidelines, patients may use a variety of email services in their everyday life, including personal email servers (hosted directly by the patient or by a mandated third party), public email providers (free services such as Gmail, Hotmail, or associated with an internet provider, eg, Free.fr or Orange in France), and business email providers (ie, emails addresses hosted by or for the patients' employer).

The use of business or personal email addresses accessed from the workplace is regulated, and could lead to privacy breaches. The US [10-13] and European [14] regulations clearly state that an employer is allowed to monitor their employees' communications under different conditions and circumstances (eg, if the employee has consented to such a monitoring or if communications are within the ordinary course of business). Therefore, patients providing their health institution with a business email put themselves at risk of privacy violation. The unsecured exchange of medical information through that channel could constitute a breach of doctor–patient confidentiality for the institution.

Data privacy can also be weakened or even shattered by different types of security breaches such as global surveillance programs (eg, the PRISM program [15] revealed by the press), computer hacking, or abuse of data sharing policies.

In this study, we used a common definition for privacy breach: unauthorized collection, use, or disclosure of personal information.

Previous studies have already described the use of email communication in health care. In a 2013 survey, Lee et al [1] showed that 37% of American patients reported having contacted their physicians via email within the last 6 months, and 18% via Facebook. In 2011, a survey across 14 European countries [4] reported that between 18.7% and 36.3% of patients sent or received at least one email from doctors, nurses, or health care organizations. Characteristics of patients using emails seem well-known. However, to the best of our knowledge the risks of privacy breaches associated with the use of different types of email providers are still to be characterized.

Study Objective

In this article, we assessed the characteristics of email addresses used by patients attending the European Hospital Georges Pompidou (HEGP) in Paris, France. We described the demographic and minimal clinical characteristics of the patients who provided an email address in contrast to those who did not. We examined what types of email services (free, personal, or business providers) were used by patients, identified the main email service providers, and studied the use of business email accounts. Finally, we discussed the potential issues associated with the type of email services.

Methods

Overview

In this section, we detail the data collection and wrangling process, the statistical methodology used to analyze the demographic characteristics, and describe the methodology used to explore email characteristics.

Study Design

The study had a cross-sectional design, and we followed the STROBE recommendation [16] for the reporting.

Study Population

Data Source

Data were extracted from the clinical information system of HEGP, a 700-bed university hospital located in Paris, France, specializing in oncology, cardiovascular disease, and emergency medicine. HEGP has been using a clinical information system since its opening in 2000 [17] and is certified on the 6th level of the Healthcare Information and Management Systems Society (HIMSS) stage [18]. An i2b2 clinical data warehouse [19] fed with data collected in the clinical information system is also deployed.

Inclusion Criteria

The study population includes all patients with a hospital stay or an outpatient visit between January 1, 2000, and November 1, 2018, dates on which data were extracted.

Data Extraction

In the remainder of the manuscript, we will refer to the part of an email address located after the character "@" as the *domain name* (ie, Local-part@domainname)

We collected the following data from the hospital's information system: the domain name of the patient's email address, sex, year of birth, year of the first contact with the hospital (stay or visit), year of last contact with the hospital, and the presence of their first or last names in the domain name of their email addresses. To preserve privacy, an algorithm automatically substituted the domain name with a random string before the extraction when the domain name included the first or last names of the patient.

Outcomes

Comparison of the Population of Patients According to the Presence of an Email Address in the Database

The presence of an email address in the database may depend on the year of first contact with the hospital. In addition, the use of email address and the administrative registration procedure are time dependent. To mitigate this risk of bias, we matched cases (patients with email address) and controls (patients without email address) on the year of first contact with the hospital using a 1:1 ratio.(univariate matching). The quality of the matching was checked (Multimedia Appendix 1). In addition to demographic data, we measured the chronic comorbidities from the ICD-10 (International Classification of Diseases, 10th revision) adaptation of the Charlson Comorbidity Index [20]. We defined the variable "at least one chronic disease" from the chronic comorbidities included in the ICD-10 phenotyping: myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, connective tissue disease, ulcer disease, mild liver disease, moderate/severe liver disease, diabetes mellitus, hemiplegia, moderate/severe renal disease, diabetes mellitus with chronic complications, any tumor, leukemia, lymphoma, metastatic solid tumor, and AIDS. A more detailed description is available in Multimedia Appendix 2.

Classification of Email Addresses

We mapped the most frequently used domain names to companies providing the email service (eg, emails ending with gmail.com, hotmail.com, orange.fr, or free.fr are associated with Google, Microsoft, Orange, and Free corporations, respectively). The companies were identified using prior knowledge and *Whois* requests when necessary. We then evaluated in the curated data the frequency of each of the providers.

We defined a business email as an email provided by someone's employer. To define business emails, we leveraged a list of free email service providers compiled by Svay [21], secondarily enriched with French and European email providers: voila.fr, dbmail.com, icloud.com, gmx.fr, bluewin.ch, dartybox.com, skynet.be, gmx.com and topnet.tn, numericable.fr, orange.fr, orange.com, wanadoo.fr, numericable.fr, outlook.fr, sfr.fr, neuf.fr, club-internet.fr, noos.fr, numericable.com, free.com, bbox.fr, live.fr. We excluded the domain names appearing in this list and then manually reviewed the remaining.

Personal emails were identified during the extraction process by the presence of the first or the last name in the domain name.

Statistical Analysis and Implementation

Data were expressed as n (%). Characteristics were compared between groups (included versus excluded population, with email address versus without email address) by Mann–Whitney *U* test for numerical data and chi-square test for categorical data. A logistic regression analysis was performed to determine characteristics associated with having an email address on the matched groups. All analysis were performed using R Statistical Software [22]. The source code and resource files (including the updated list of email providers) are available online [23].

Ethical Considerations

The study was authorized by the HEGP Institutional Review Board on December 21, 2018. The HEGP Institutional Review Board [24] has waived the requirement to obtain informed consent or specific approval for this study.

Results

Population

Between 2000 and August of 2018, we collected 41,004 patient email addresses (4.22% of the total population of patients, ie, 41,004/971,822). Of the total number of patients in the database (971,822 patients), we selected a subgroup of 41,004 patients as a control group (ie, patients who did not provide an email address). The characteristics of the population and the control group are described in Table 1. Statistical tests are presented in Multimedia Appendix 3.

Table 1. Characteristics of the patients (before and after matching).

Characteristics	Overall population (before matching; N=971,822)	With email (before matching; N=41,004)	Population (after matching; N=82,008)	Excluded population of matching (N=889,814)
Male, n (%)	485,737 (49.98)	19,463 (47.47)	39,957 (48.72)	445,780 (50.10)
Year of birth, median (Q1-Q3)	1960 (1944-1978)	1964 (1952-1978)	1965 (1951-1981)	1960 (1943-1977)
Year of birth, n (%)				
Before 1950	342,676 (35.26)	9308 (22.70)	20,069 (24.47)	322,607 (36.26)
1950-1969	282,381 (29.06)	16,115 (39.30)	27,735 (33.82)	254,646 (28.62)
1970-1989	281,390 (28.95)	12,655 (30.86)	26,022 (31.73)	255,368 (28.70)
1990 and more	65,375 (6.73)	2926 (7.14)	8182 (9.98)	57,193 (6.43)
Year of the first contact, median (Q1-Q3)	2009 (2004-2014)	2015 (2011-2017)	2015 (2011-2017)	2008 (2003-2013)
Year of the first contact, n (%)				
2005 and before	283,848 (29.21)	3875 (9.45)	7750 (9.45)	276,098 (31.03)
2005-2010	229,823 (23.65)	4198 (10.24)	8396 (10.24)	221,427 (24.88)
2010 and after	458,151 (47.14)	32,931 (80.31)	65,862 (80.31)	392,289 (44.09)
Year of last contact, median (Q1-Q3)	2012 (2006-2016)	2018 (2016-2018)	2017 (2014-2018)	2011 (2005-2015)
Year of last contact, n (%)				
2010 and before	283,848 (29.21)	468 (1.14)	7375 (8.99)	414,718 (46.61)
2011-2014	229,823 (23.65)	4198 (10.24)	14,109 (17.20)	215,483 (24.22)
2015-2018	458,151 (47.14)	35,609 (86.84)	60,459 (73.72)	259,613 (29.18)
At least one chronic disease, n (%)	8072 (0.83)	5828 (14.21)	8072 (9.84)	0[a]

[a]ICD-10 (International Classification of Diseases, 10th revision) codes were not included in the electronic heath record before 2008.

Description of the Email Service Providers

The 10 most frequent email providers and their frequency on the data set are summarized in Table 2.

Table 2. Prevalence of the 10 most frequent email providers used by HEGP patients.

Email type and provider	Description (N=41,004)	Headquarters localization
Business, n	2878	
Personal, n	347	
Email provider, n (%)	37,779	
Google	11,730 (28.61)	USA
Microsoft	7802 (19.03)	USA
Orange	6402 (15.61)	Europe
Yahoo!	5071 (12.37)	USA
Free	2427 (5.92)	Europe
SFR	1170 (2.85)	Europe
La Poste	815 (1.99)	Europe
AOL	330 (0.80)	USA
NOOS	319 (0.78)	Europe
Bouygues Telecom	250 (0.61)	Europe

Identification of Business Emails

Starting from 41,004 email addresses, we identified 37,779 email addresses associated with known email providers. We removed 347 emails likely to use personalized domain names (ie, those that included the first or last name of the patient) and identified 2878 domain names likely to be associated with business email addresses.

For privacy reasons, we did not detail the nature of the companies present in the list. However, we can highlight notable findings: around 100 people used email addresses from our own institution. We also found more than 50 email addresses associated with other governmental institutions.

Unsurprisingly, among the most represented companies, businesses geographically located close to the hospital are on top of the list.

Matching Case–Control: Comparison of the Population of Patients According to the Presence of an Email Address

Table 3 presents the results of the comparison of the groups with and without an email address. After matching, we observed a significant difference in gender, year of birth, and follow-up (P<.001). The proportion of patients born between 1950 and 1969 with an email address was higher (39.30 [16,115/41,004] versus 28.34% [11,620/41,004]), whereas we observed the opposite for patients born after 1989 (7.14% [2926/41,004] versus 12.82% [5256/41,004]). Table 4 presents the results of the logistic regressions.

Table 3. Case–control comparison of groups with and without emails.

Comparison	Overall population (N=82,008)	Without email address (N=41,004)	With email address (N=41,004)	*P*
Male, n (%)	39,957 (48.72)	20,494 (49.98)	19,463 (47.47)	<.001
Year of birth, median (Q1-Q3)	1965 (1951-1981)	1967 (1950-1984)	1964 (1952-1978)	<.001
Year of birth, n (%)				<.001
Before 1950	20,069 (24.47)	10,761 (26.24)	9308 (22.70)	
1950-1969	27,735 (33.82)	11,620 (28.34)	16,115 (39.30)	
1970-1989	26,022 (31.73)	13,367 (32.60)	12,655 (30.86)	
1990 and after	8182 (9.98)	5256 (12.82)	2926 (7.14)	
Year of last contact, median (Q1-Q3)	2017 (2014-2018)	2016 (2013-2017)	2018 (2016-2018)	<.001
Year of last contact, n (%)				<.001
2010 and before	7375 (8.99)	6907 (16.84)	468 (1.14)	
2011-2014	14,109 (17.20)	9247 (22.55)	4862 (11.86)	
2015-2018	60,459 (73.72)	24,850 (60.60)	35,609 (86.84)	
At least one chronic disease, n (%)	8072 (9.84)	2244 (5.47)	5828 (14.21)	<.001

Table 4. Regressions (univariate and multivariate analyses) in the matched groups (N=82,008).

Characteristics	Value	Univariate analysis	Multivariate analysis
		Crude odds ratio (95% CI)	Adjusted odds ratio (95% CI)
Male, n (%)	39,957 (48.72)	0.90 (0.88-0.93)	0.85 (0.83-0.88)
Year of birth, n (%)			
Before 1950	20,069 (24.47)	Reference	Reference
1950-1969	27,735 (33.82)	1.60 (1.55-1.66)	1.60 (1.54-1.67)
1970-1989	26,022 (31.73)	1.09 (1.05-1.14)	1.17 (1.12-1.22)
After 1990	8182 (9.98)	0.64 (0.61-0.68)	0.56 (0.53-0.59)
Year of last contact, n (%)			
2010 and before	7375 (8.99)	Reference	Reference
2011-2014	14,109 (17.20)	7.76 (7.03-8.58)	7.66 (6.93-8.47)
2015-2018	60,459 (73.72)	21.1 (19.3-23.3)	20.8 (18.9-22.9)
At least one chronic disease, n (%)	8072 (9.84)	2.86 (2.72-3.01)	2.23 (2.11-2.36)

Discussion

Principal Results

Between 2000 and August 2018, out of 971,822 recorded patients, HEGP patients have provided 41,004 email addresses. Patients providing email addresses were mostly born between 1950 and 1989 (28,770/41,004, 70.16%) and had their first contact (ie, first visit) with the hospital after 2010. Among patients who provided an address, the follow-up duration was mostly less than 3 months (11,309/41,004, 27.58%), or more than 60 months (10,847/41,004, 26.45%).

Among all the email addresses collected (N=41,004), 37,779 corresponded to known email providers. Among the top 10 email providers (covering 36,316/37,779, 96.12%, of known domain names), 65.99% (24,933/37,779) were hosted by companies with headquarters located in the USA, versus 30.13% (11,383/37,779) hosted by companies located in Europe; 7.02% (2878/41,004) of the patients provided business email addresses, and 0.85% (347/41,004) used a personalized domain name.

Compared with patients who did not provide an email address, patients who provided one were older. Table 4 shows that patients suffering from at least one chronic disease were also more likely to provide an email address (odds ratio [OR] 2.23, 95% CI 2.11-2.36).

The systematic collection of email addresses in our institution was put in place only recently, explaining the strong effect of the year of last contact (OR 20.8, 95% CI 18.9-22.9).

We counterintuitively observed a lower proportion of email addresses in younger patients. A selection bias in the hospital population is likely to be blamed. Indeed, our hospital is specializing in cardiovascular diseases, oncology, and emergency medicine. Younger patients mostly treated for nonchronic diseases or visiting the emergency department are less likely to provide their address than older patients with a regular follow-up for a chronic condition in specialized departments.

Comparison With Prior Work

To the best of our knowledge no institution has provided a description of email providers used by their patients. Newhouse et al [4] have described characteristics of patients using email for health care communication. They showed that the proportion of patients using emails increased with the number of health problems reported, the presence of a current long-standing illness or health problem, the presence of undergoing long-term medical treatment, and the number of visits to the doctor in the last 12 months. These observations are in line with our results. However, they reported a proportion of 18.70% of the French population using emails for health care communication (against 4.22% [41,004/971,822] for our hospital). Newhouse et al [4] also reported a more important use of email for male and younger patients. The difference with our results could be explained by the population studied and a bias of declaration. Their results come from a survey in the general population, whereas our results are reflecting a hospital population. Furthermore, they are based on the declaration of emails by the patients themselves at the hospital registration service. Our results are not comparable with the US survey of Lee et al [1], in which no difference between gender was observed; however, the age distribution is comparable to our population. The study population of the Lee et al [1] survey came from retail pharmacy users.

The comparison with related studies should be made carefully as the French and US populations are not strictly comparable; furthermore, the organization of the French and US health systems are substantially different.

Limitations

This study has some limitations. The first is a selection bias which may limit the generalizability of its findings. This is a cross-sectional study with a large period of inclusion but our population comes from a university hospital in Paris and is not representative of the national population.

This study only reflects the official collection of email addresses. It is possible that health care professionals use direct email communication with their patients outside of the hospital information system. Therefore, the volume of email addresses collected might be underestimated. However, it is likely that the distribution of domain names remains similar.

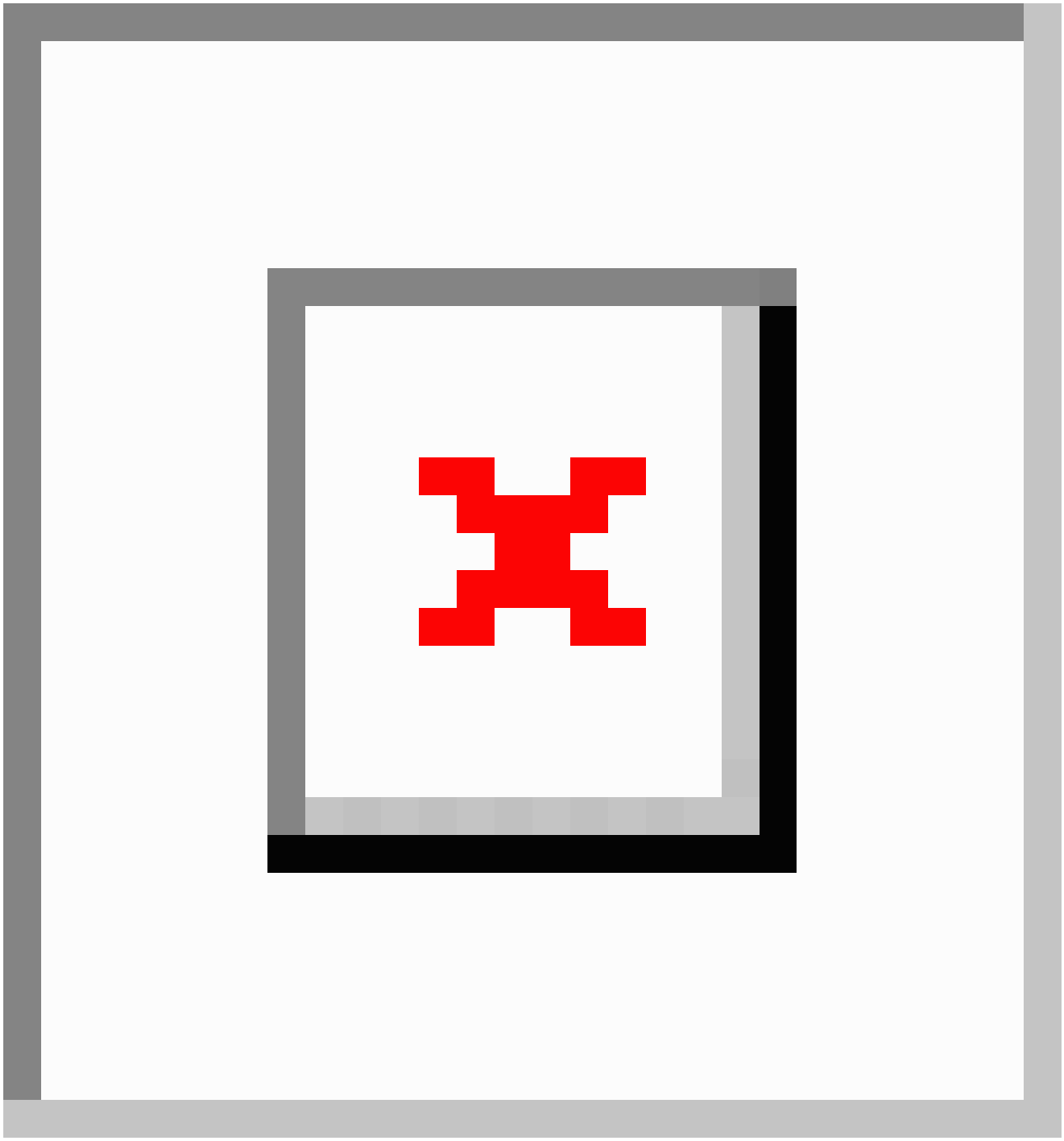
Another limitation is a bias in the measurement of chronic comorbidities from the ICD-10 according to [20]. Between 2000 and 2007, our information system did not include ICD-10 codes in the electronic heath records in routine care. Nevertheless, this bias is nondifferential after our matching on the first year of contact.

Perspectives and Ethical Considerations

Privacy Breach: Analysis of Risks

Table 4 shows that the patients more likely to have provided the institution with their email address are those suffering from chronic diseases. Social discrimination may be more acute for patient suffering from chronic diseases than for other pathologies. Moreover, these patients are more solicited by hospitals and health care professionals for different reasons: access to results, invitations to participate in clinical studies, and information regarding reuse of the data in retrospective studies. We explored the risks associated with such communication. To determine the risk, we analyzed the likelihood of such an issue and its consequences (Figure 1).

Figure 1. Risk analysis of unauthorized access to private email communication between a hospital and a patient.



The main risk lies in the interception of the email by a third party. Two main cases appear very probable: (1) breach of privacy through smartphone notifications, given that only a limited number of smartphone users are able to fine-tune the privacy settings of their smartphones and tablets; (2) breach of privacy through family members: the email address provided in the information system may be used by other members of the family. It could be a shared email address with their spouse, the address of a child for elderly patients, or that of a parent for younger patients. The shared address collected at a point in time might not reflect the patients' aspirations for privacy at a later date.

The hacking of email accounts and information accessed by employers are both likely. It is worth noting that business emails can legally be monitored. In such a case the infrastructure to read and capture the information may already be deployed. In the case of email hacking, additional steps are needed.

Many health professionals now use encrypted emails to exchange information with their colleagues. In such cases, the privacy of the exchange is guaranteed by technological means. However, most of the secured email solutions in health care are designed to protect communications among health care professionals, and not with patients. The direct communication via emails may put the patients at risk of a privacy breach.

Strategies for Safe Communication

In the communication between health institutions and patients, 3 strategies can be identified: a push strategy (from the

institution to the patient, eg, sending messages and reminders), a pull strategy (from the institution to secured third parties, eg, collecting patients' information from booking platforms), or a bidirectional strategy. The first 2 are already taken into account with the current communication protocols by most of the health care providers. However, unsolicited direct communication from the patients to their physicians can lead to privacy breaches. Health organizations increasingly rely on secured platforms to transmit health documents to their patients (eg, laboratory results, hospital web-based patient portals [25]). However, the communication between the physicians and their patients need to be included in the exchange solutions. There is here a strong need for secured and user-friendly communication tools for direct communication between patients and physicians.

Note that in the context of health care, the name of the institution can reveal sensitive information (eg, an email from the ABC Oncology Center, or the Infection Disease Department at ABC could inform regarding the pathology of the patient). To mitigate this risk, the use of an email address from a general care provider (eg, ABC hospitals) might be preferable.

The Self-Privacy Harms

Patients can expose their health data to harm when they use unsecured communication channels (and especially emails). Regulators engaged the responsibility of the institution through the GDPR and HIPAA guidance (in Europe and the United States), but cannot fully protect citizens against their own data sharing behavior.

Protection of data privacy is based on both accountable health care professionals and *engaged* citizen-patients (informed, implicated, and responsible). The education of citizens regarding data privacy and governance should be promoted and encouraged.

The demand, and legitimate need, for easier ways of communication might overlap strongly with the privacy paradox. First mentioned by Brown in 2001 [26] and defined by Norberg et al [27] in 2007, the privacy paradox is "the relationship between individuals' intentions to disclose personal information and their actual personal information disclosure behaviors." It reveals an inconsistency of privacy attitudes and privacy behavior [28]. The use of cash-back rewards cards or free internet services represent famous examples of the privacy paradox. In both cases, the monetary benefits for the user are often inferior to the price the user would assign to the data provided in exchange. In the case of free email providers, virtually all providers are bound by the law to respect and enforce privacy protection.

Acknowledgments

BR was partially funded by the CARPEM research program (SIRIC CARPEM).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Quality check of the case-control matching of groups with and without emails.
[DOCX File , 13 KB - jmir_v23i2e13992_app1.docx]

Multimedia Appendix 2
Chronic disease comparison of case-control groups.
[DOCX File , 15 KB - jmir_v23i2e13992_app2.docx]

Multimedia Appendix 3
Comparison between the population after matching and the excluded population.
[DOCX File , 15 KB - jmir_v23i2e13992_app3.docx]

References

1. Lee JL, Choudhry NK, Wu AW, Matlin OS, Brennan TA, Shrank WH. Patient Use of Email, Facebook, and Physician Websites to Communicate with Physicians: A National Online Survey of Retail Pharmacy Users. J Gen Intern Med 2016 Jan;31(1):45-51 [FREE Full text] [doi: 10.1007/s11606-015-3374-7] [Medline: 26105675]
2. Brooks RG, Menachemi N. Physicians' use of email with patients: factors influencing electronic communication and adherence to best practices. J Med Internet Res 2006 Mar 24;8(1):e2 [FREE Full text] [doi: 10.2196/jmir.8.1.e2] [Medline: 16585026]
3. Singh H, Fox SA, Petersen NJ, Shethia A, Street RL. Older patients' enthusiasm to use electronic mail to communicate with their physicians: cross-sectional survey. J Med Internet Res 2009 Jun 16;11(2):e18 [FREE Full text] [doi: 10.2196/jmir.1143] [Medline: 19632972]

4. Newhouse N, Lupiáñez-Villanueva F, Codagnone C, Atherton H. Patient use of email for health care communication purposes across 14 European countries: an analysis of users according to demographic and health-related factors. J Med Internet Res 2015 Mar 06;17(3):e58 [FREE Full text] [doi: 10.2196/jmir.3700] [Medline: 25798912]
5. Prensky M. Digital Natives, Digital Immigrants Part 1. On the Horizon 2001 Sep;9(5):1-6. [doi: 10.1108/10748120110424816]
6. Baynes C. Hackers steal 1.5 million people's personal data in cyber attack on Singapore's health service. 2018. URL: https://www.independent.co.uk/news/world/asia/singapore-personal-data-hack-health-database-government-millions-people-a8456111.html [accessed 2021-02-01]
7. Confessore N. Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. New York Times. 2018. URL: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html [accessed 2021-02-01]
8. Alder S. HIPAA Compliance Guide. HIPAA Journal. 2017. URL: https://www.hipaajournal.com/hipaa-compliance-guide/ [accessed 2021-02-01]
9. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing. Directive 95/46/EC.: Off J Eur Union; 2016. URL: http://data.europa.eu/eli/reg/2016/679/oj [accessed 2021-02-07]
10. United States v. Axselle, 604 F.2d 1330, 1338 (10th Cir. 1979). 1979. URL: https://www.leagle.com/decision/19791934604f2d133011720 [accessed 2021-02-07]
11. United States v. Amen, 831 F.2d 373, 378 (2d Cir. 1987). 1987. URL: https://www.leagle.com/decision/19871204831f2d37311123 [accessed 2021-02-07]
12. Arias v. Mutual Central Alarm Service, Inc., 202 F.3d 553, 557 n.3 (2d Cir. 2000). 2000. URL: https://www.leagle.com/decision/2000755202f3d5531697 [accessed 2021-02-07]
13. United States v. Mullins, supra note 16, 992 F.2d,1472 (1993). 1993. URL: https://www.leagle.com/decision/19932464992f2d147212180 [accessed 2021-02-07]
14. Grand Chamber judgment of 12 January 2016 in the case of Barbulescu v. Romania (no. 61496/08). European Court of Human Rights. URL: https://www.echr.coe.int/Pages/home.aspx?p=hearings&w=6149608_30112016&language=en&c=&py=2017 [accessed 2021-02-07]
15. Gellman B, Poitras L. U. S. URL: https://www.washingtonpost.com/investigations/us-intelligence-mining-data-from-nine-us-internet-companies-in-broad-secret-program/2013/06/06/3a0c0da8-cebf-11e2-8845-d970ccb04497_story.html [accessed 2021-02-07]
16. von EE, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007 Oct 20;370(9596):1453-1457. [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
17. Degoulet P, Marin L, Lavril M, Le Bozec C, Delbecke E, Meaux J, et al. The HEGP component-based clinical information system. Int J Med Inform 2003 Mar;69(2-3):115-126. [doi: 10.1016/s1386-5056(02)00101-6] [Medline: 12810117]
18. URL: http://www.himssanalytics.org/amam [accessed 2021-02-07]
19. Zapletal E, Rodon N, Grabar N, Degoulet P. Methodology of integration of a clinical data warehouse with a clinical information system: the HEGP case. Stud Health Technol Inform Internet 2010:2010.
20. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. Journal of Clinical Epidemiology 2004 Dec;57(12):1288-1294. [doi: 10.1016/j.jclinepi.2004.03.012]
21. free-email-providers Internet. Svay M. URL: https://github.com/mauricesvay/free-email-providers/blob/master/free-email.txt [accessed 2013-05-28]
22. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Core team, R Foundation for Statistical Computing; 2015.
23. Looten V, Rance B. EmailsPrivacy. URL: https://github.com/equipe22/EmailsPrivacy [accessed 2021-02-01]
24. Jannot A, Zapletal E, Avillach P, Mamzer M, Burgun A, Degoulet P. The Georges Pompidou University Hospital Clinical Data Warehouse: A 8-years follow-up experience. International Journal of Medical Informatics 2017 Jun;102:21-28. [doi: 10.1016/j.ijmedinf.2017.02.006]
25. Jahn M, Porter B, Patel H, Zillich A, Simon S, Russ A. Usability Assessment of Secure Messaging for Clinical Document Sharing between Health Care Providers and Patients. Appl Clin Inform 2018 Jun 27;09(02):467-477. [doi: 10.1055/s-0038-1660521]
26. Brown B. Studying the internet experience. HP Laboratories Technical Report (HPL--49) Internet. 2001. URL: http://www.hpl.hp.com/techreports/2001/HPL-2001-49.pdf [accessed 2021-02-07]
27. Norberg P, Horne D, Horne D. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. J Consum Aff Internet 2007 Jun; 2007 Jun;41(1):126. [doi: 10.1111/j.1745-6606.2006.00070.x]
28. Kokolakis S. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. Computers & Security 2017 Jan;64:122-134. [doi: 10.1016/j.cose.2015.07.002]

Abbreviations

GDPR: General Data Protection Regulation
HEGP: European Hospital Georges Pompidou
HIMSS: Healthcare Information and Management Systems Society
HIPAA: Health Insurance Portability and Accountability Act
ICD-10: International Classification of Diseases, 10th revision
OR: odds ratio

Edited by G Eysenbach; submitted 12.03.19; peer-reviewed by R Haase, D Jessup, M Reynolds, J Farzi; comments to author 03.10.19; revised version received 20.07.20; accepted 10.08.20; published 24.02.21.

Please cite as:
Looten V, Neuraz A, Garcelon N, Burgun A, Chatellier G, Rance B
The Epidemiology of Patients' Email Addresses in a French University Hospital: Case-Control Study
J Med Internet Res 2021;23(2):e13992
URL: https://www.jmir.org/2021/2/e13992
doi:10.2196/13992
PMID:33625375

Original Paper

Hidden Variables in Deep Learning Digital Pathology and Their Potential to Cause Batch Effects: Prediction Model Study

Max Schmitt[1*], MSc; Roman Christoph Maron[1*], MSc; Achim Hekler[1], MSc; Albrecht Stenzinger[2], MD; Axel Hauschild[3], MD; Michael Weichenthal[3], MD; Markus Tiemann[4], MD; Dieter Krahl[5], MD; Heinz Kutzner[6], MD; Jochen Sven Utikal[7,8], MD; Sebastian Haferkamp[9], MD; Jakob Nikolas Kather[10], MD; Frederick Klauschen[11], MD; Eva Krieghoff-Henning[1], PhD; Stefan Fröhling[12], MD; Christof von Kalle[13], MD; Titus Josef Brinker[1], MD

[1]Digital Biomarkers for Oncology Group, National Center for Tumor Diseases, German Cancer Research Center (DKFZ), Heidelberg, Germany
[2]Institute of Pathology, University Hospital Heidelberg, University of Heidelberg, Heidelberg, Germany
[3]Department of Dermatology, University Hospital Kiel, University of Kiel, Kiel, Germany
[4]Institute for Hematopathology Hamburg, Hamburg, Germany
[5]Private Institute of Dermatopathology, Heidelberg, Germany
[6]Private Institute of Dermatopathology, Friedrichshafen, Germany
[7]Skin Cancer Unit, German Cancer Research Center (DKFZ), Heidelberg, Germany
[8]Department of Dermatology, University Medical Center Mannheim, University of Heidelberg, Mannheim, Germany
[9]Department of Dermatology, University Hospital of Regensburg, Regensburg, Germany
[10]Department of Medicine III, RWTH University Hospital Aachen, Aachen, Germany
[11]Institute of Pathology, Charité University Hospital Berlin, Berlin, Germany
[12]National Center for Tumor Diseases, German Cancer Research Center (DKFZ), Heidelberg, Germany
[13]Department of Clinical-Translational Sciences, Charité and Berlin Institute of Health, Berlin, Germany
[*]these authors contributed equally

Corresponding Author:
Titus Josef Brinker, MD
Digital Biomarkers for Oncology Group
National Center for Tumor Diseases
German Cancer Research Center (DKFZ)
Im Neuenheimer Feld 460
Heidelberg, 69120
Germany
Phone: 49 6221 3219304
Email: titus.brinker@dkfz.de

Abstract

Background: An increasing number of studies within digital pathology show the potential of artificial intelligence (AI) to diagnose cancer using histological whole slide images, which requires large and diverse data sets. While diversification may result in more generalizable AI-based systems, it can also introduce hidden variables. If neural networks are able to distinguish/learn hidden variables, these variables can introduce batch effects that compromise the accuracy of classification systems.

Objective: The objective of the study was to analyze the learnability of an exemplary selection of hidden variables (patient age, slide preparation date, slide origin, and scanner type) that are commonly found in whole slide image data sets in digital pathology and could create batch effects.

Methods: We trained four separate convolutional neural networks (CNNs) to learn four variables using a data set of digitized whole slide melanoma images from five different institutes. For robustness, each CNN training and evaluation run was repeated multiple times, and a variable was only considered learnable if the lower bound of the 95% confidence interval of its mean balanced accuracy was above 50.0%.

Results: A mean balanced accuracy above 50.0% was achieved for all four tasks, even when considering the lower bound of the 95% confidence interval. Performance between tasks showed wide variation, ranging from 56.1% (slide preparation date) to 100% (slide origin).

Conclusions: Because all of the analyzed hidden variables are learnable, they have the potential to create batch effects in dermatopathology data sets, which negatively affect AI-based classification systems. Practitioners should be aware of these and similar pitfalls when developing and evaluating such systems and address these and potentially other batch effect variables in their data sets through sufficient data set stratification.

(J Med Internet Res 2021;23(2):e23436) doi:10.2196/23436

KEYWORDS

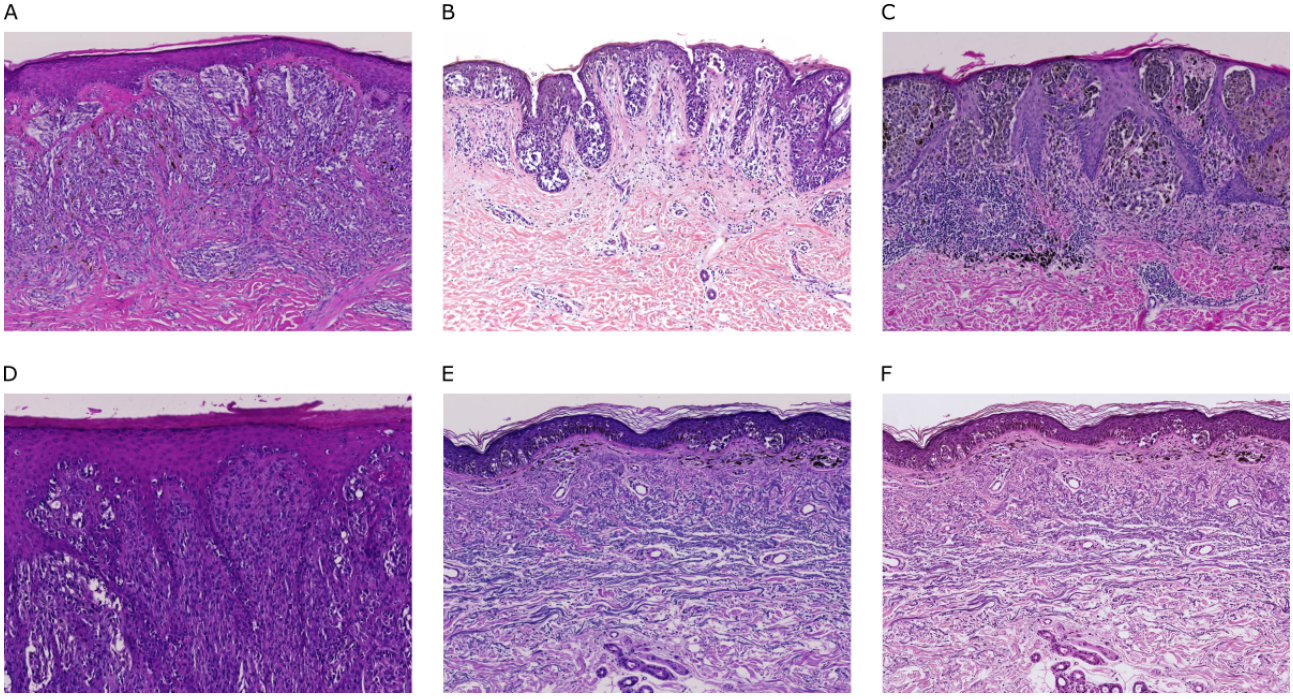
artificial intelligence; machine learning; deep learning; neural networks; convolutional neural networks; pathology; clinical pathology; digital pathology; pitfalls; artifacts

Introduction

The advent of artificial intelligence (AI) in digital pathology (DP) has resulted in the development of various algorithms for the detection, classification, and further evaluation of multiple cancer subtypes [1]. General interest and enthusiasm for this emerging technology continues to grow, exemplified by the development of a variety of convolutional neural network (CNN)–based oncology systems for the analysis of histological images of breast [2,3], lung [4], skin [5,6], and gastrointestinal [7] cancer. However, the successful implementation of CNN-based assistance systems in DP is complicated by a plethora of challenges [8-10], some of which are domain-specific, while others are omnipresent in the field of deep learning (DL) and machine learning (ML) in general.

One important issue in the field of biomedical data analysis is the occurrence of batch effects, which are defined as differences among subsets of a data set introduced through technological artifacts [11,12]. In DP, such artifacts are introduced during tissue processing and slide preparation [13], and presumably also during slide digitization, image compression, and storage [14], all of which affect slide and image appearance (Figure 1). We expand on this definition of batch effects by including biological factors, presumably unrelated to the actual classification task, as causative agents. Both factors (biological and nonbiological) are referred to as hidden variables from here on.

Figure 1. Comparison of exemplary whole slide image sections obtained at different institutes: (A) Institute of Pathology, University Hospital Heidelberg, University of Heidelberg, Heidelberg, Germany (Zeiss scanner; Carl Zeiss AG); (B) Department of Dermatology, University Hospital Kiel, University of Kiel, Kiel, Germany (3DHISTECH scanner; 3DHISTECH Ltd); (C) Private Institute of Dermatopathology, Mönchhofstraße 52, Heidelberg, Germany (Zeiss scanner); (D) Department of Dermatology, University Medical Center Mannheim, University of Heidelberg, Mannheim, Germany (Zeiss scanner); and (E) Private Institute of Dermatopathology, Siemensstraße 6/1, Friedrichshafen, Germany (Zeiss scanner). (F) The same slide section as shown in (E) but scanned with a Hamamatsu scanner (Hamamatsu Photonics KK) rather than a Zeiss scanner.



Batch effects can be problematic during development of ML models, where hidden variables are learned instead of or in addition to the intended target variables. The hidden variable then acts as a complete or partial proxy for the intended target variable, negatively affecting the model's performance. Studies have addressed this issue by focusing on various normalization techniques [15-19]. In addition, standardized preprocessing procedures and balanced data set construction may aid in reducing but not eliminating batch effects. Overall, this is concerning, and a previous study using a breast cancer tissue cohort has already suggested the existence of batch effects in

parts of the publicly available Cancer Genome Atlas (TCGA) pathology repository [20].

We expand on these findings by analyzing the learnability of four exemplary selected hidden variables, as learnable variables can cause batch effects that can negatively influence DL algorithms. Our research on hidden variables aims to highlight that batch effects are not an unlikely occurrence, thereby reinforcing the importance of proper data set construction and experimental design, as well as sensitizing the community toward these and similar pitfalls for the emerging field of DL in DP.

Methods

Study Design

Using a proprietary dermatopathological data set of anonymized slides, a series of ML tasks were formulated, where each task investigated the learnability of a certain hidden variable. The analyzed variables are believed to be found throughout DP whole slide image (WSI) data sets, but their learnability does not necessarily need to generalize.

Next, multiple DL models of the same architecture were trained on the variable stated by each task, followed by subsequent performance analysis, where an assessment was made of whether the task's variable was learnable or unlearnable. A variable was considered learnable when the 95% confidence interval of its mean balanced accuracy had a lower bound above 50.0% when calculated on the slide level. Note that a random classifier, which is unable to learn the variable, would be expected to achieve a balanced accuracy of approximately 50.0%.

Ethics approval was obtained from the ethics committee of the Medical Faculty of Mannheim, University of Heidelberg, Mannheim, Germany.

Data Set

A total of 427 hematoxylin and eosin–stained preparations were obtained from five different institutes, with each slide belonging to one patient and containing tissue sections of melanoma biopsies (Table 1). For details on the slide digitization process, see Multimedia Appendix 1.

Table 1. Overview of individual data sets.

Data set	Origin of slides	Number of slides	Number of tiles	Scanner type	Tasks
1	Heidelberg[a]	81	1,344,825	Zeiss[b]	2 (slide preparation date); 3 (slide origin)
2	Kiel[c]	196	2,092,726	3DHISTECH[d]	1 (patient age); 2 (slide preparation date)
3	Heidelberg[e]	73	832,940	Zeiss	3 (slide origin)
4	Friedrichshafen[f]	54	350,518; 364,196	Zeiss; Hamamatsu[g]	3 (slide origin); 4 (scanner type)
5	Mannheim[h]	23	513,256	Zeiss	3 (slide origin)

[a]Institute of Pathology, University Hospital Heidelberg, University of Heidelberg, Heidelberg, Germany.

[b]Carl Zeiss AG.

[c]Department of Dermatology, University Hospital Kiel, University of Kiel, Kiel, Germany.

[d]3DHISTECH Ltd.

[e]Private Institute of Dermatopathology, Mönchhofstraße 52, Heidelberg, Germany.

[f]Private Institute of Dermatopathology, Siemensstraße 6/1, Friedrichshafen, Germany.

[g]Hamamatsu Photonics KK.

[h]Department of Dermatology, University Medical Center Mannheim, University of Heidelberg, Mannheim, Germany.

Classification Tasks

Four classification tasks were performed, each analyzing one predefined hidden variable. Data sets for each task were chosen based on data availability, while simultaneously minimizing the risk of cross-task learning. For instance, all data sets were used for the slide origin prediction task (task 3) except data set 2, as the Department of Dermatology, University Hospital Kiel, used a different scanner type (3DHISTECH scanner; 3DHISTECH Ltd) to digitize the slides (Table 1). Therefore, a classifier could potentially determine the slide origin for data set 2 by determining the scanner type.

Task 1: Patient Age

To determine patient age, data set 2 was used, and only slides with an assigned patient age were analyzed. Slides were divided into one of two classes based on patient age (≤48 years versus >78 years), excluding slides of patients with ages in between. The cutoff points were chosen in an effort to achieve a natural balance between both age groups, with the 30-year gap making it plausible to observe possible distinct age-dependent morphological features.

Task 2: Slide Preparation Date

To determine the year of slide preparation, data sets 1 and 2 were used. Data availability varied but was generally sufficient for years 2014-2018. For each data set, separate binary classification tasks were defined, where slides were taken from every other year to ensure that there was a minimum of 365 days between the preparation dates of slides from each class (eg, data set 1, 2015 versus 2017). This resulted in five separate classification subtasks.

Task 3: Slide Origin

To determine the origin of the respective slide, all data sets except data set 2 were used. Origin was defined as the institution from which the slides were obtained.

Task 4: Scanner Type

To determine the scanner type, data set 4 was used. These slides were scanned twice, but due to the slight difference in resolution between the Zeiss (Carl Zeiss AG) and Hamamatsu (Hamamatsu Photonics KK) scanners, scanned Zeiss slides were reprocessed specifically for this task by downscaling their resolution (0.22 µm/px) to match the resolution of the Hamamatsu scanner (0.23 µm/px).

Model Training

Each task had a designated combined data set based on the setup outlined above. Each combined data set was divided into a training set and a test set on slide level using an 80:20 split. If the resulting test set contained fewer than 10 slides, a cross-validation approach on slide level was employed to increase the size of the test set to a minimum of 10 slides per class.

A ResNet50 architecture was trained for each task. In cases where cross-validation was used for testing, the number of trained CNNs equaled the number of cross-validation folds. The training run for each task was repeated a total of five times to obtain a robust average performance uninfluenced by stochastic training events. This number was chosen arbitrarily but with the intention to reduce the overall computing time. For exact technical details on the cross-validation and training procedure, please see Multimedia Appendix 1 or refer to Multimedia Appendix 2 for an exemplary jupyter notebook demonstrating the basic training procedure.

Model Inference and Statistical Evaluation

Inference was carried out on each task's respective test set using the complete set of tiles for each slide. The class for a WSI was computed by first predicting on its complete set of tiles, averaging all output probabilities, and assigning the class label with the highest average probability to the slide. Because each training and evaluation run for a task was repeated five times, a mean balanced accuracy with a corresponding 95% confidence interval could be computed.

Results

Learnability was investigated on the slide level, as that is the standard and decisive criterion in DP. For all tasks, balanced accuracy was generally higher on slide level than on tile level.

For each task, a balanced accuracy over 50.0% was achieved, even when taking into account the range of the corresponding confidence intervals. Classifier performance varied widely inter-task and intra-task for task 2, which had multiple subtasks. Task 1 (patient age) had a mean balanced accuracy of 87.5% (Table 2). For task 2 (slide preparation date), performance varied widely between subtasks, ranging from 56.1% (95% CI 52.7% to 59.5%) to 83.5% (95% CI 80.9% to 86.1%). Classifiers for task 3 (slide origin) and 4 (scanner type) showed balanced accuracies of 97.9% (95% CI 97.3% to 98.5%) and 100%, respectively.

Table 2. Overall mean performance of each task's classifiers measured using mean balanced accuracy and evaluated on tile level and slide level.

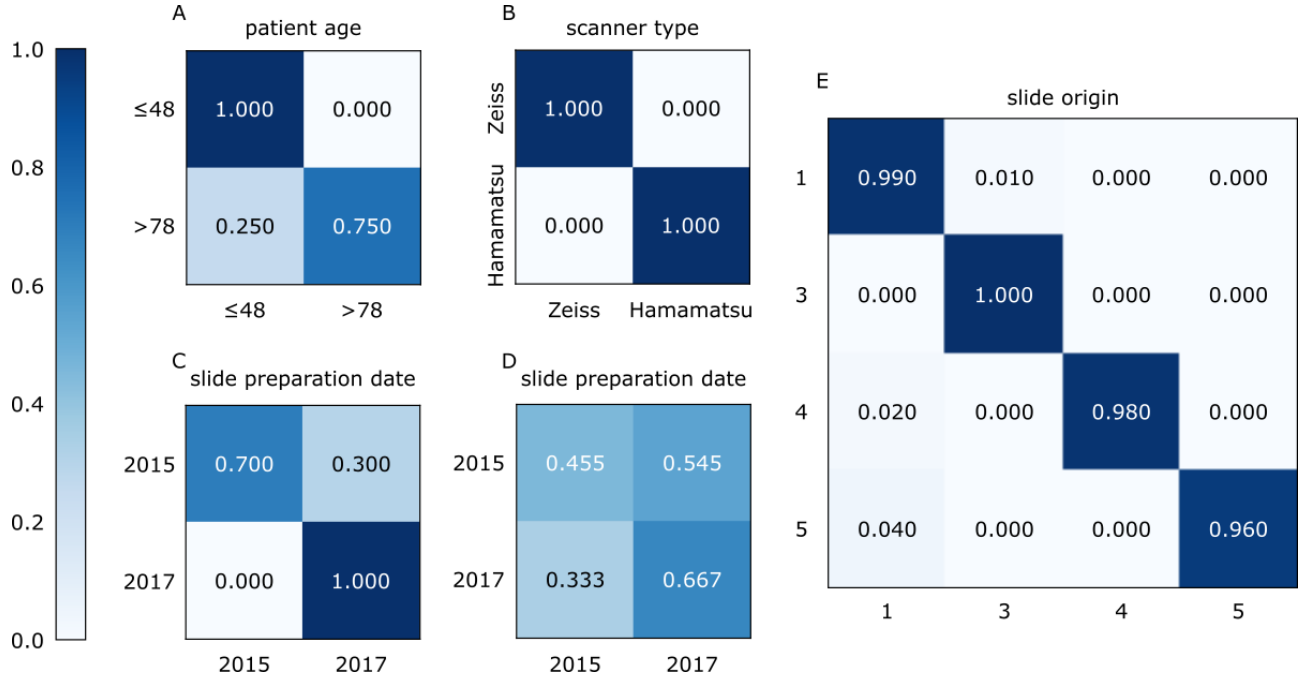
	ResNet50 performance (mean balanced accuracy)	
Task[a]	Tile level	Slide level (95% CI)[b]
1: Patient age	76.2%	87.5%
2: Slide preparation date		
Data set 1: 2015 versus 2017	54.1%	56.1% (52.7% to 59.5%)
Data set 1: 2016 versus 2018	56.5%	63.2% (53.4% to 73.0%)
Data set 2: 2014 versus 2016	69.0%	82.0% (76.4% to 87.6%)
Data set 2: 2015 versus 2017	66.6%	83.5% (80.9% to 86.1%)
Data set 2: 2016 versus 2018	52.7%	56.7% (52.6% to 60.7%)
3: Slide origin	94.2%	97.9% (97.3% to 98.5%)
4: Scanner type	100%	100%

[a]Test sets for each task had a minimum of 10 slides per class.

[b]Confidence intervals are shown for the decisive criteria (slide level) and are omitted for tasks where no variation on slide level was observed.

Looking at the distributions of each task-specific model for the first run, slide origin and scanner type could be predicted with very high accuracy, with minor misclassification errors for task 3 (slide origin) and no misclassifications for task 4 (scanner type). For task 1, patient age below 48 years could be predicted with high accuracy, but one-quarter of the slides originating from older patients (>78 years) were erroneously classified as belonging to the younger age group. For task 2 (slide preparation date), the results varied widely between comparisons. The 2-year comparison with the highest balanced accuracy (data set 2, 2015 versus 2017) showed some misclassifications, with slides from 2015 occasionally being classified as 2017, whereas the task with the lowest balanced accuracy (data set 1, 2015 versus 2017) showed frequent misclassifications in both directions (Figure 2).

Figure 2. Distributions of the models' predictions for tasks 1 to 4 on slide level for ResNet50 (run 1). (A) Task 1 (patient age) prediction. (B) Task 4 (scanner type) prediction. (C, D) Comparison of the two most distinct classification subtasks for task 2 (slide preparation date), where balanced accuracy was either at a maximum (C: data set 2, 2015 versus 2017) or at a minimum (D: data set 1, 2015 versus 2017). (E) Task 3 (slide origin) prediction, with the data set number displayed rather than the data set origin for better readability.



Discussion

Principle Findings

Using four exemplary hidden variables found in DP WSI data sets, we showed that these variables were learned by a DL algorithm for a dermatopathological data set. A learnable hidden variable may cause a batch effect, which can greatly affect the training of such algorithms if said variable is unintendedly picked up instead of or in addition to the intended target variable. We hypothesized that there would be hidden variables that would figure very prominently on the slides. These are more likely to be picked up and used by the algorithm to classify the slides and therefore pose the greatest threat to classification accuracy. To identify such "high risk" variables, we limited the amount of training the CNN received, using a standard architecture and a training procedure with little optimization and no training until convergence. This should result in only prominent variables being learned, which likely pose the greatest threat to a classifier's accuracy, although an influence of factors that may be learned by a more extensive training procedure cannot be excluded.

All of the four variables tested (patient age, slide preparation date, slide origin, and scanner type) were learned by the classifier, albeit to different extents. The highest balanced accuracy was observed for task 4 (scanner type). For this task, slides from data set 4 were scanned twice with different devices but processed with the same image-processing pipeline, leaving differences in scanner type (eg, specific scanner hardware or software) as the only causative source of variations, which must be quite pronounced based on the high balanced accuracy.

Performances for tasks 3 (slide origin) and 4 (scanner type) were comparable. As all slides for task 3 were scanned and processed using the same scanner type and pipeline, digitization as the source of the observed batch effects can be ruled out. Therefore, the source of variation most likely stemmed from the slide preparation step, a complex process with lots of potential variables related to the sectioning, fixation, staining, and mounting procedures. Determining the origin of images was previously shown outside the field of DP, where a CNN correctly identified hospital systems based solely on chest radiographs [21].

The aforementioned slide preparation step was also a likely contributor to the "classifiability" of the slide preparation dates (task 2), together with slide aging itself (ie, small amounts of tissue and dye degradation). While slide aging is presumably a gradual process, the slide preparation step is expected to change more abruptly when institutions change the exact mode of slide preparation over time (eg, when a new staining protocol is introduced). This could explain the large variability in classifier performance for task 2, where in some cases the difference between two particular years could be identified much more accurately than between two other years at the same institution. Based on these results, the learnability of this task depends highly on the chosen year and data set, making the underlying variable less of a risk factor but still worthwhile considering.

The patient age prediction task was the only task that reflected a true "biological" difference between the analyzed groups. It is known that the texture of the skin changes as result of the aging process via a multitude of processes [22]. For instance, the amount of elastin and collagen decreases with time, which results in a restructuring of the fibrous tissue in deeper skin layers. On UV-exposed skin, additional similar effects may also be induced by photoaging. Moreover, some extent of biological variability regarding the aging process likely exists. This rather complex pattern of skin aging may explain why the ability of

the CNN to separate the chosen age groups was not perfect. Performance would likely drop if the age gap of 30 years were to be reduced or if instead of WSIs showing both tumor and healthy skin, only tumor regions were regarded. Nonetheless, the classifier's ability to pick up variables representing large age differences is important to consider, as an unequal patient age distribution is not unlikely to occur for certain medical DL objectives, especially since cancer incidence increases with age.

Based on these findings, it is not unlikely that the discussed variables may interfere with the generation of an accurate CNN-based classifier. Due to the technical nature of slide preparation date, slide origin, and scanner type, their learnability could generalize to other fields in DP, which may have to be investigated in further studies. The learnability of patient age, however, may be more specific to the field of dermatology. While patient age is known to have an impact on the skin, age-related differences may be much less prominent in other tissues.

Prevention and Verification

In order to minimize the learning of batch effect variables, we suggest to balance any known batch effect variables during creation of the training data set, in addition to any normalization and preprocessing standardization. If easy-to-learn variables are equally balanced between classes, separation based on these variables should no longer result in a reduction of the training loss, thus losing its optimization value. In addition, large and diversified validation sets decrease the likelihood that unwanted correlations between batch effect variables and the intended biological variable exist and thereby can aid in uncovering whether the intended biological variable or some confounding hidden variable was learned.

Unfortunately, balancing training data sets for all potential batch effects is unfeasible. Even randomized clinical trials can only ever be balanced for a few features that are deemed crucial to the comparison in question. With time, additional knowledge accumulating both within and outside the field of AI-supported medicine may help researchers to clarify which potential influencing variables have to be taken into account for which tasks. In this regard, the future realization of more transparent AI-systems that facilitate both explainability and causability [23] would go a long way in helping practitioners to better assess the reliability of AI-systems through a better understanding of their decision-making process.

Limitations

A major limitation is that the list of considered artifacts is not exhaustive. As numerous other potential confounders exist, some of which we are not yet aware of, complete coverage of all possible artifacts in one study is impossible and therefore has to be limited to a selection considered to be crucial. However, because of the black-box nature of DL algorithms, there is no proof to show what the model actually learned, meaning that any of the unaccounted for artifacts could by chance correlate with a task's class distribution and be learned instead. This cannot be ruled out; however, by increasing validation set size and diversity, chances of these accidental training set correlations persisting through to the validation set decrease and should therefore be detected during the validation stage.

In this study, only results obtained with one DL architecture are shown. We therefore investigated two additional architectures (DenseNet121 and VGG16) and observed a similar trend (see Multimedia Appendix 1).

While learnability was only investigated on a dermatopathological data set in this study, some of the insights gained here may be transferable to other fields in DP. Moreover, this study did not intend to show exactly what variables can be learned but rather to show that unexpected variables can be learned.

Conclusions

Our DL model was able to learn several potential batch effect variables with relative ease, a finding that is likely to be applicable to other DL models, too. Thus, our results highlight the importance of identifying and minimizing these effects, not only by normalization and preprocessing standardization but also by carefully constructing training and validation sets for DL classification tasks.

Acknowledgments

This study was funded by the Federal Ministry of Health, Berlin, Germany (grant: Skin Classification Project; grant holder: Dr Titus J Brinker, Digital Biomarkers for Oncology Group, National Center for Tumor Diseases, German Cancer Research Center). The sponsor had no role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Conflicts of Interest

A Hauschild receives clinical trial support, speaker's honoraria, or consultancy fees from the following companies: Amgen, Bristol Myers Squibb, Merck Serono, MSD, Novartis, Oncosec, Philogen, Pierre Fabre, Provectus, Regeneron, Roche, OncoSec, Sanofi-Genzyme, and Sun Pharma. Outside of the submitted work, JSU is on the advisory board or has received honoraria and travel support from Amgen, Bristol Myers Squibb, GlaxoSmithKline, LeoPharma, Merck Sharp and Dohme, Novartis, Pierre Fabre, and Roche. Outside of the submitted work, SH has advisory roles for or has received honoraria from Pierre Fabre Pharmaceuticals, Novartis, Roche, Bristol Myers Squibb, Amgen, and MSD. TJB owns a company that develops mobile apps (Smart Health Heidelberg GmbH, Handschuhsheimer Landstr. 9/1, 69120 Heidelberg).

Multimedia Appendix 1
Additional methods and results.
[DOCX File , 26 KB - jmir_v23i2e23436_app1.docx]

Multimedia Appendix 2
Exemplary jupyter notebook.
[ZIP File (Zip Archive), 6 KB - jmir_v23i2e23436_app2.zip]

References

1. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 2019 Nov;16(11):703-715 [FREE Full text] [doi: 10.1038/s41571-019-0252-y] [Medline: 31399699]
2. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, the CAMELYON16 Consortium, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA 2017 Dec 12;318(22):2199-2210 [FREE Full text] [doi: 10.1001/jama.2017.14585] [Medline: 29234806]
3. Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NNC, et al. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. Sci Rep 2017 Apr 18;7:46450 [FREE Full text] [doi: 10.1038/srep46450] [Medline: 28418027]
4. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med 2018 Oct;24(10):1559-1567. [doi: 10.1038/s41591-018-0177-5] [Medline: 30224757]
5. Hekler A, Utikal JS, Enk AH, Berking C, Klode J, Schadendorf D, et al. Pathologist-level classification of histopathological melanoma images with deep neural networks. Eur J Cancer 2019 Jul;115:79-83 [FREE Full text] [doi: 10.1016/j.ejca.2019.04.021] [Medline: 31129383]
6. Hekler A, Utikal JS, Enk AH, Solass W, Schmitt M, Klode J, et al. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. Eur J Cancer 2019 Sep;118:91-96 [FREE Full text] [doi: 10.1016/j.ejca.2019.06.012] [Medline: 31325876]
7. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med 2019 Jul;25(7):1054-1056. [doi: 10.1038/s41591-019-0462-y] [Medline: 31160815]
8. Higgins C. Applications and challenges of digital pathology and whole slide imaging. Biotech Histochem 2015 Jul;90(5):341-347. [doi: 10.3109/10520295.2015.1044566] [Medline: 25978139]
9. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. Med Image Anal 2016 Oct;33:170-175 [FREE Full text] [doi: 10.1016/j.media.2016.06.037] [Medline: 27423409]
10. Tizhoosh HR, Pantanowitz L. Artificial Intelligence and Digital Pathology: Challenges and Opportunities. J Pathol Inform 2018;9:38 [FREE Full text] [doi: 10.4103/jpi.jpi_53_18] [Medline: 30607305]
11. Goh WWB, Wang W, Wong L. Why Batch Effects Matter in Omics Data, and How to Avoid Them. Trends Biotechnol 2017 Jun;35(6):498-507. [doi: 10.1016/j.tibtech.2017.02.012] [Medline: 28351613]
12. Phan JH, Quo CF, Cheng C, Wang MD. Multiscale integration of -omic, imaging, and clinical data in biomedical informatics. IEEE Rev Biomed Eng 2012;5:74-87 [FREE Full text] [doi: 10.1109/RBME.2012.2212427] [Medline: 23231990]
13. Taqi SA, Sami SA, Sami LB, Zaki SA. A review of artifacts in histopathology. J Oral Maxillofac Pathol 2018;22(2):279 [FREE Full text] [doi: 10.4103/jomfp.JOMFP_125_15] [Medline: 30158787]
14. Janowczyk A, Basavanhally A, Madabhushi A. Stain Normalization using Sparse AutoEncoders (StaNoSA): Application to digital pathology. Comput Med Imaging Graph 2017 Apr;57:50-61 [FREE Full text] [doi: 10.1016/j.compmedimag.2016.05.003] [Medline: 27373749]
15. Kothari S, Phan JH, Stokes TH, Osunkoya AO, Young AN, Wang MD. Removing batch effects from histopathological images for enhanced cancer diagnosis. IEEE J Biomed Health Inform 2014 May;18(3):765-772 [FREE Full text] [doi: 10.1109/JBHI.2013.2276766] [Medline: 24808220]
16. Kothari S, Phan JH, Wang MD. Scale normalization of histopathological images for batch invariant cancer diagnostic models. Annu Int Conf IEEE Eng Med Biol Soc 2012;2012:4406-4409 [FREE Full text] [doi: 10.1109/EMBC.2012.6346943] [Medline: 23366904]
17. Roy S, Kumar Jain A, Lal S, Kini J. A study about color normalization methods for histopathology images. Micron 2018 Nov;114:42-61. [doi: 10.1016/j.micron.2018.07.005] [Medline: 30096632]
18. Van Eycke Y, Allard J, Salmon I, Debeir O, Decaestecker C. Image processing in digital pathology: an opportunity to solve inter-batch variability of immunohistochemical staining. Sci Rep 2017 Feb 21;7:42964 [FREE Full text] [doi: 10.1038/srep42964] [Medline: 28220842]

19. Kothari S, Phan J, Moffitt R, Stokes T, Hassberger S, Chaudry Q, et al. Automatic batch-invariant color segmentation of histological cancer images. Proc IEEE Int Symp Biomed Imaging 2011;2011:657-660 [FREE Full text] [doi: 10.1109/ISBI.2011.5872492] [Medline: 27532016]
20. Janowczyk A, Zuo R, Gilmore H, Feldman M, Madabhushi A. HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides. JCO Clin Cancer Inform 2019 Apr;3:1-7 [FREE Full text] [doi: 10.1200/CCI.18.00157] [Medline: 30990737]
21. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLoS Med 2018 Nov;15(11):e1002683 [FREE Full text] [doi: 10.1371/journal.pmed.1002683] [Medline: 30399157]
22. Rittié L, Fisher GJ. Natural and sun-induced aging of human skin. Cold Spring Harb Perspect Med 2015 Jan 05;5(1):a015370 [FREE Full text] [doi: 10.1101/cshperspect.a015370] [Medline: 25561721]
23. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 2019;9(4):e1312 [FREE Full text] [doi: 10.1002/widm.1312] [Medline: 32089788]

Abbreviations

AI: artificial intelligence
CNN: convolutional neural network
DL: deep learning
DP: digital pathology
ML: machine learning
TCGA: The Cancer Genome Atlas
WSI: whole slide image

Edited by G Eysenbach; submitted 12.08.20; peer-reviewed by A Böer-Auer, JA Benítez-Andrades, A Holzinger; comments to author 10.10.20; revised version received 14.10.20; accepted 06.12.20; published 02.02.21.

Please cite as:
Schmitt M, Maron RC, Hekler A, Stenzinger A, Hauschild A, Weichenthal M, Tiemann M, Krahl D, Kutzner H, Utikal JS, Haferkamp S, Kather JN, Klauschen F, Krieghoff-Henning E, Fröhling S, von Kalle C, Brinker TJ
Hidden Variables in Deep Learning Digital Pathology and Their Potential to Cause Batch Effects: Prediction Model Study
J Med Internet Res 2021;23(2):e23436
URL: https://www.jmir.org/2021/2/e23436
doi:10.2196/23436
PMID:33528370

Original Paper

Use and Control of Artificial Intelligence in Patients Across the Medical Workflow: Single-Center Questionnaire Study of Patient Perspectives

Simon Lennartz[1], MD; Thomas Dratsch[1], DPhil, MD; David Zopfs[1], MD; Thorsten Persigehl[1], MD; David Maintz[1], MD; Nils Große Hokamp[1], MD; Daniel Pinto dos Santos[1], MD

Institute for Diagnostic and Interventional Radiology, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

Corresponding Author:
Daniel Pinto dos Santos, MD
Institute for Diagnostic and Interventional Radiology
Faculty of Medicine and University Hospital Cologne
University of Cologne
Kerpener Straße 62
Cologne, 50937
Germany
Phone: 49 22147896063
Fax: 49 22147882384
Email: daniel.pinto-dos-santos@uk-koeln.de

Abstract

Background: Artificial intelligence (AI) is gaining increasing importance in many medical specialties, yet data on patients' opinions on the use of AI in medicine are scarce.

Objective: This study aimed to investigate patients' opinions on the use of AI in different aspects of the medical workflow and the level of control and supervision under which they would deem the application of AI in medicine acceptable.

Methods: Patients scheduled for computed tomography or magnetic resonance imaging voluntarily participated in an anonymized questionnaire between February 10, 2020, and May 24, 2020. Patient information, confidence in physicians vs AI in different clinical tasks, opinions on the control of AI, preference in cases of disagreement between AI and physicians, and acceptance of the use of AI for diagnosing and treating diseases of different severity were recorded.

Results: In total, 229 patients participated. Patients favored physicians over AI for all clinical tasks except for treatment planning based on current scientific evidence. In case of disagreement between physicians and AI regarding diagnosis and treatment planning, most patients preferred the physician's opinion to AI (96.2% [153/159] vs 3.8% [6/159] and 94.8% [146/154] vs 5.2% [8/154], respectively; *P*=.001). AI supervised by a physician was considered more acceptable than AI without physician supervision at diagnosis (confidence rating 3.90 [SD 1.20] vs 1.64 [SD 1.03], respectively; *P*=.001) and therapy (3.77 [SD 1.18] vs 1.57 [SD 0.96], respectively; *P*=.001).

Conclusions: Patients favored physicians over AI in most clinical tasks and strongly preferred an application of AI with physician supervision. However, patients acknowledged that AI could help physicians integrate the most recent scientific evidence into medical care. Application of AI in medicine should be disclosed and controlled to protect patient interests and meet ethical standards.

(J Med Internet Res 2021;23(2):e24221) doi:10.2196/24221

KEYWORDS

artificial intelligence; clinical implementation; questionnaire; survey

Introduction

The incremental use of artificial intelligence (AI) is widely considered as one of the most disruptive developments of the past decades [1]. In medicine, increasing evidence has revealed the promising applications of AI for disease prevention, diagnosis, and treatment [2-4]. Among these, specialties relying on the interpretation of medical imaging data, such as

dermatology [5], pathology [6], and radiology [7-9], have a particular scientific and economic focus.

While several studies suggested that AI might outperform board-certified physicians at narrow diagnostic tasks [10-13], broad clinical implementation of such technologies has not matched the pace of scientific advancements. Among practical reasons, such as high heterogeneity in clinical data and clinical workflows as well as cost efficiency considerations affecting clinical implementation, other concerns pertain to ethical questions and liability. Consequently, different concepts of AI implementation in existing clinical workflows in a controlled and responsible manner have been discussed to preserve pivotal pillars of accountability [14]. In this regard, one important consensus is that AI should remain as transparent and explainable as possible.

To implement AI in an acceptable manner, it is important to understand perspectives on AI use in clinical routines from among stakeholders, including patients and health care professionals such as students, physicians, and caregivers [15]. In this respect, most studies have reported that medical professionals in specialties most evidently influenced by AI agree with its implementation and consider it a tool complementing the armamentarium they regularly work with. Despite dire early predictions—for example, radiologists being potentially replaced by AI [16]—recent studies have demonstrated a more gradual adaptation of AI solutions that currently augment human capabilities rather than replacing them [17].

Regarding patients' perspectives on the use of AI in medicine, current studies are primarily focused on particular subspecialties or individual diagnostic procedures [18-20]. Although most of these studies indicate that patients generally accept the implementation of AI in medicine, a recent study investigating patients' perspectives on implementing specific AI devices revealed controversial opinions among numerous patients, particularly regarding the question of human control [21].

Therefore, this study aimed to investigate patients' perspectives on the clinical implementation of AI in a more coherent approach that includes the assessment of key clinical competencies such as physician–patient interaction, diagnosis, and therapy as integral parts of the medical workflow. Further, we investigated opinions for human control of AI and its acceptance depending on different disease severities.

Methods

Participants

After our single-center survey study was approved by the institutional review board (approval number 19-1552), patients scheduled for cross-sectional imaging between February 10, 2020, and May 24, 2020, were informed about the possibility to voluntarily complete an anonymous questionnaire on registration. Dedicated boxes for returning the completed questionnaires were placed in the waiting areas. Questionnaires were collected and data were manually transferred to a structured spreadsheet (Excel, Microsoft Corp) at the end of the acquisition period. The response rate was calculated as proposed by the American Association for Public Opinion Research, using the following formula:

where C is the number of completed questionnaires; P, the number of partially completed questionnaires; and R, the number of nonresponders not consenting to participate or opting out of the questionnaire study.

Questionnaire

A senior expert in medical AI, a radiology resident with 4 years of experience and a PhD in psychology, conducted a literature review on previous surveys on AI in general and in medicine. As previous surveys were focused on particular subspecialties or limited in their scope, a new survey was generated. Patient interaction, diagnostics, and treatment decisions were identified as key clinical competencies and were therefore implemented as central elements to assess patients' acceptance towards the application of AI in medicine. Because no previous questionnaire comprising all the specific endpoints of our study was available, external validation was omitted. The questionnaire comprised five subsections. The first subsection inquired age, gender, level of education, and a history of a cancer diagnosis. Moreover, prior knowledge of AI had to be indicated. In the second subsection, participants were asked about their confidence in physicians versus AI in different clinical tasks, including the assessment of the medical history of a patient, making of diagnostic and treatment decisions, and addressing of the patients' fears and need for information. The third subsection determined patients' opinions on human control of AI at diagnosis and treatment planning. In the fourth subsection, the respondents were asked to state whose decision should be preferred at diagnosis and treatment planning in case of disagreement between the physician and AI. Finally, in the fifth subsection, participants were asked to indicate their acceptance regarding the application of AI in diagnosing and treating diseases of different severity. In the second, third, and fifth subsections, patients were asked to indicate their agreement based on a Likert scale ranging from 1 ("I strongly disagree") to 5 ("I strongly agree"). In subsection 4, a binary decision between AI and physician was requested, with the option to choose "I don't know/I don't have an opinion on this."

Sample Size Estimation

A power analysis was performed for sample size estimation. In order not to disregard smaller, yet important differences in this new research field, our study was sufficiently powered to detect small effects (Cohen d=0.2). With α=.05 and power=0.80, the projected sample size needed to detect a small effect (Cohen d=0.2) for within-group comparisons was N=199 [22]. Therefore, the inclusion of at least 200 participants was deemed necessary.

Statistical Analysis

Data were analyzed using SPSS (version 25, IBM Corp) as well as R 3.4.0 with RStudio 1.0.136 [23]. Likert scores were compared using two-tailed, paired samples t tests. Chi-squared tests were used to compare the proportions of participants. Pearson correlation analysis was used to assess the association

between prior knowledge of AI and confidence on physicians and AI. A *P* value below .05 was considered statistically significant.

Results

Participants

Questionnaire outcomes obtained from 229 patients (99 male, 112 female, 18 of unspecified gender; age 18-82 years) scheduled for computed tomography (CT) or magnetic resonance imaging (MRI) were included. The demographic characteristics, education, prior knowledge of AI and the history of cancer diagnosis of the questionnaire participants are summarized in Table 1. In total, 515 questionnaires were handed out, of which 229 were completed and 19 were incomplete (response rate 48.2%).

Table 1. Data obtained from the questionnaire participants (N=229).

Characteristic	Participants
Age (years), mean (SD)	51.8 (15.4)
Gender, n (%)	
Men	99 (43.2)
Women	112 (48.9)
Nonbinary	0 (0)
Not indicated	18 (7.9)
Level of education, n (%)	
1 (ISCED[a] 1-2)	57 (24.9)
2 (ISCED 3-5)	71 (31.0)
3 (ISCED 6-8)	84 (36.7)
Not indicated	17 (7.4)
Prior knowledge of artificial intelligence (1=completely unfamiliar; 5=very familiar), n (%)	
1	39 (17)
2	46 (20.1)
3	79 (34.5)
4	28 (12.2)
5	8 (3.5)
Not indicated	29 (12.7)
History of oncologic disease, n (%)	
Previous cancer diagnosis	144 (62.9)
No previous cancer diagnosis	66 (28.8)
Not indicated	19 (8.3)

[a]ISCED: International Standard Classification of Education. ISCED levels: 1=primary education, 2=lower secondary education, 3=upper secondary education, 4=postsecondary nontertiary education, 5=short-cycle tertiary education, 6=bachelor or equivalent, 7=master or equivalent, 8=doctoral or equivalent.

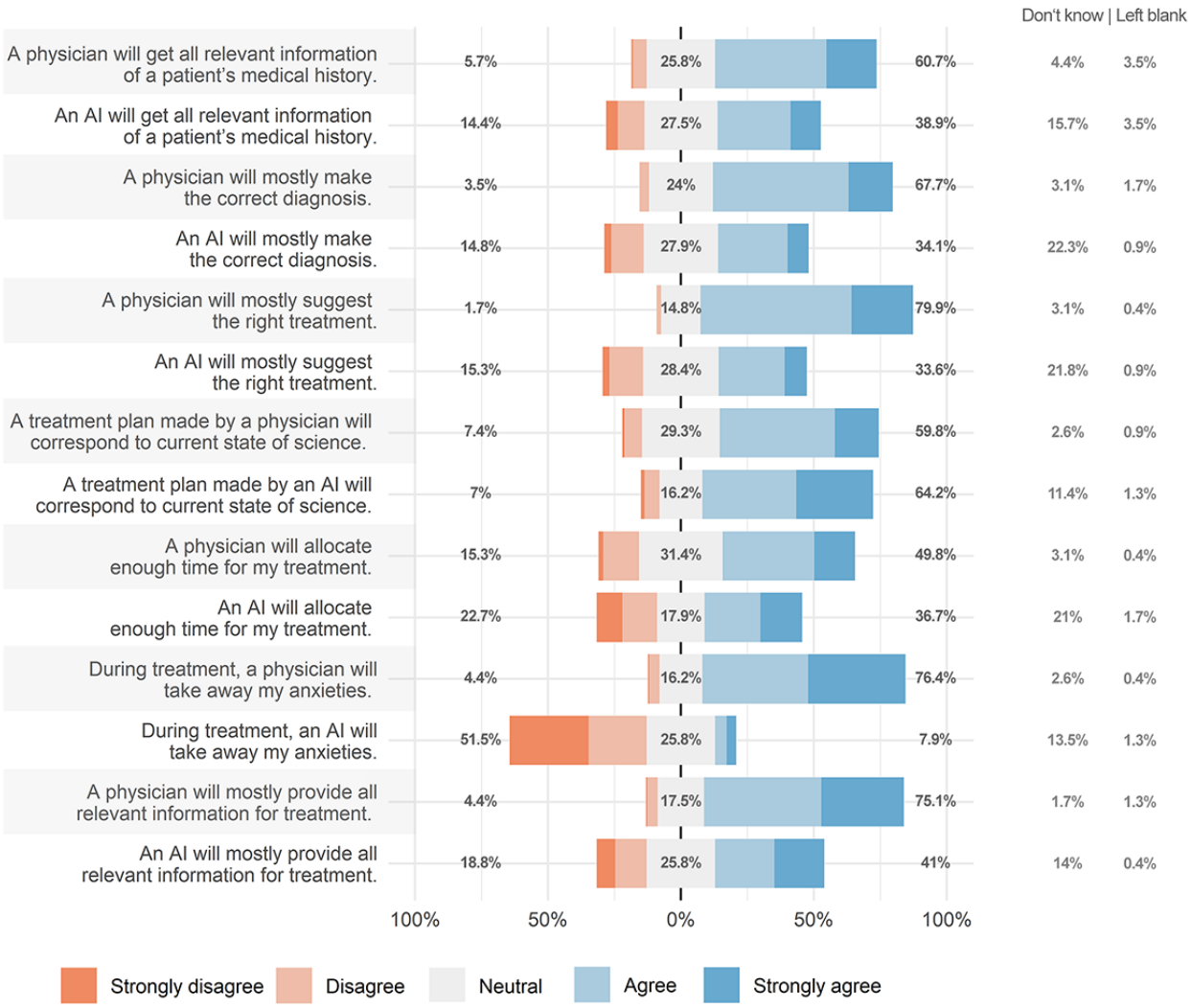
Confidence in the Capabilities of Physicians vs AI

Patients assigned significantly higher mean scores to the physician rather than to AI for all capabilities included (Table 2), except for treatment planning based on the most recent scientific evidence, for which the participants favored AI to physicians (3.96 [SD 0.95] vs 3.71 [SD 0.84]; mean difference –0.255; 95% CI –0.416 to –0.094; t_{195}=–3.12; *P*=.002, Cohen *d*=–0.233). Figure 1 summarizes the proportions of ratings assigned to physicians and AI for different clinical tasks.

Table 2. Comparison of mean ratings regarding confidence in clinical capabilities of physicians and artificial intelligence (AI).

Capability	Physician, mean (SD)	AI, mean (SD)	Mean difference (95% CI)	*t* (*df*)	*P* value	Cohen *d*
Obtaining any relevant information from my medical history	3.81 (0.85)	3.38 (1.04)	0.433 (0.258 to 0.608)	4.88 (177)	<.001	0.407
Making an accurate diagnosis	3.91 (0.76)	3.27 (0.96)	0.633 (0.454 to 0.812)	6.99 (168)	<.001	0.617
Proposing the appropriate treatment	4.08 (0.64)	3.30 (0.98)	0.78 (0.621 to 0.94)	9.67 (172)	<.001	0.959
Planning treatment according to recent state of science	3.71 (0.84)	3.96 (0.95)	–0.255 (–0.416 to –0.094)	–3.12 (195)	.002	0.233
Allocating a sufficient amount of time for me	3.55 (0.99)	3.25 (1.30)	0.295 (0.043 to 0.547)	2.31 (172)	.02	0.208
Taking away my worries and addressing my anxieties	4.15 (0.86)	2.16 (1.08)	1.984 (1.79 to 2.179)	20.16 (192)	<.001	1.653
Providing all information relevant to my treatment	4.08 (0.80)	3.38 (1.20)	0.699 (0.488 to 0.911)	6.52 (192)	<.001	0.598

Figure 1. Results of the questionnaire regarding the clinical capabilities of physicians versus artificial intelligence (AI). Percentages refer to the proportion of negative (light orange, orange), neutral (gray), and positive (light blue, blue) responses. Proportions of patients who indicated "Don't know" or left the question blank are indicated on the right. Patients favored physicians to AI for all clinical capabilities except for making a treatment plan based on current clinical knowledge, where they preferred an AI algorithm.



Different Levels of Human Control of AI at Diagnosis and Treatment Planning

During both diagnosis and treatment, patients were significantly more comfortable with the use of AI under the physician's supervision than without such supervision (at diagnosis: 3.90 [SD 1.20] vs 1.64 [SD 1.03]; mean difference 2.26; 95% CI 2.08 to 2.43; t_{213}=25.19; *P*=.001; Cohen *d*=1.62; treatment planning: 3.77 [SD 1.18] vs 1.57 [SD 0.96]; mean difference=2.20; 95% CI 2.03 to 2.38; t_{209}=25.12; *P*=.001; Cohen *d*=1.58).

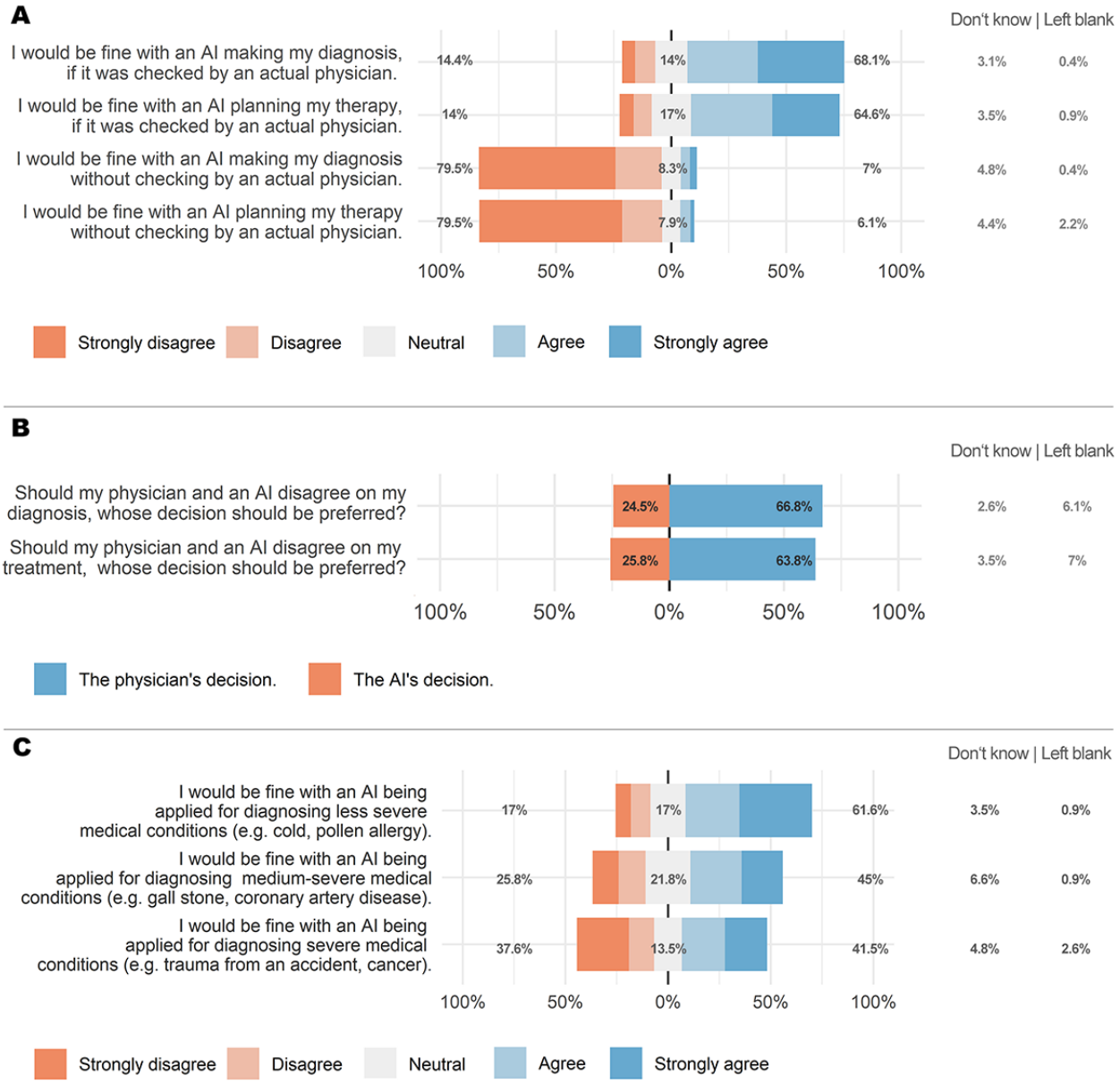
Disagreement Between Physicians and AI

When asked whose decision should be followed in case of disagreement between the physician and AI regarding the diagnosis, 66.8% (153/229) of patients decided that the diagnosis of the physician should be followed, 2.6% (n=6) of patients decided that the decision of the AI should be followed, 24.5% (n=56) of patients responded that they were undecided, and 6.1% (n=14) of patients did not respond to the question. When analyzing the responses of patients who decided on either AI or a physician, a significantly larger proportion of patients (153/159, 96.2%) decided that the diagnosis of the physician should be considered (χ^2_1=135.91; *P*=.001). The same applied to disagreement regarding treatment decisions, for which a similarly large proportion of participants (146/154, 94.8%) decided that the treatment suggested by the physician should be considered (χ^2_1=123.66; *P*=.001).

Application of AI for the Diagnosis and Treatment of Diseases of Different Severity

There was a significant main effect for disease severity ($F_{2,414}$=51.75; *P*=.001; η^2_p=0.200), indicating that the acceptance of AI was lower for more severe diseases than for less severe diseases. Post hoc *t* tests revealed that the acceptance of AI was significantly lower for diseases of medium severity (3.29 [SD 1.32]) than for those of low severity (3.77 [SD 1.27]; mean difference=0.48; 95% CI 0.33 to 0.64; t_{211}=6.15; *P*=.001; Cohen *d*=0.426). Additionally, the acceptance of AI was significantly lower for diseases of high severity (2.97 [SD 1.52]) than for diseases of medium severity (3.30 [SD 1.33]; mean difference=0.33; 95% CI 0.23 to 0.43; t_{207}=6.42; *P*=.001; Cohen *d*=0.498). Figure 2 provides an overview of the proportions of ratings regarding human control of AI, disagreement between AI and physicians, and acceptance of AI for the diagnosis and treatment of diseases of different severity.

Figure 2. Results of the questionnaire regarding control of artificial intelligence (AI) (A), disagreement between AI and physicians (B), and the application of AI for diagnosing medical conditions of low, medium, and high severity (C). Percentages refer to the proportion of negative (light orange, orange), neutral (gray), and positive (light blue, blue) responses. Proportions of patients who indicated "Don't know" or left the question blank are indicated on the right.



Correlation Between Patient-Related Factors and Responses

Prior knowledge of AI was by far the most important patient-related factor correlating with certain patients' opinions on AI. As shown in Table 3, prior knowledge of AI was significantly correlated with the acceptance of AI in almost all questionnaire items, indicating that patients assigning a higher rating to their prior knowledge on AI were generally more accepting of the use of AI for various aspects of medical treatment and diagnosis. However, the strength of the correlation was weak overall.

Table 3. Correlation between prior knowledge of artificial intelligence (AI) and assigned ratings for clinical capabilities of physicians and AI.

Criteria	Correlation between prior knowledge of AI and assigned ratings to physicians		Correlation between prior knowledge of AI and assigned ratings to AI	
	r	*P* value	*r*	*P* value
Physician/AI would be capable of obtaining any relevant information from my medical history	–0.015	.83	.193	.01
Physician/AI would be capable of making an accurate diagnosis	0.112	.12	0.170	.03
Physician/AI would be capable of proposing the appropriate treatment	0.103	.16	0.213	.007
Physician/AI would be capable of planning my treatment according to recent state of science	0.062	.39	0.295	<.001
Physician/AI would be capable of allocating a sufficient amount of time for me	0.012	.87	0.269	<.001
Physician/AI would be capable of taking away my worries and addressing my anxieties	–0.001	.99	0.310	<.001
Physician/AI would be capable of providing all information relevant to my treatment	0.016	.83	0.206	.007

Discussion

Over the past few years, research and development in AI has gained considerable momentum [24]. Although AI in medicine is faced with critical challenges regarding its implementation and reimbursement [25], AI technologies will increasingly impact the diagnosis, management, and treatment of diseases in the future.

This study shows that patients would trust physicians over AI in most clinical capabilities except for basing treatment decisions on the most current clinical knowledge, for which AI was considered superior. This is an interesting finding, as it may be increasingly difficult for physicians both in academia and private practice to keep up with the rapidly growing literature in their corresponding subspecialties [26]. Our results indicate that patients seem to be aware of this issue and consider AI superior in incorporating the most recent scientific evidence. It is worth noting that the discrepancy between the acceptance of AI and physicians was largest in the category of "taking away my worries and addressing my anxieties," for which physicians received the highest ratings. While this is, to an extent, an expected result, it clearly underlines the demand for empathetic doctor-patient interactions, concurrent with previous findings [27].

Interestingly, patients favored the capabilities of physicians to AI for diagnosis and treatment decisions, although these two aspects have received extensive media coverage, showing promise in its application in medicine. However, it is important to note that although physicians received higher ratings, a large proportion of patients still had positive views on the clinical capabilities of AI. This "cautious optimism" regarding the usefulness of AI in medicine was reflected by the broad consensus among patients for the use of AI for diagnosing and treating rather mild medical conditions; however, this optimism significantly declined with an increase in disease severity. This is in line with a previous survey outlining that acceptance of AI-based decisions declines when the stakes or risks are higher [28]. This observation is relevant in view of many commercially available algorithms for diagnosing life-threatening disease conditions such as cerebral hemorrhage, pulmonary embolism, or pneumothorax [29-31].

Because we are in the era of "narrow AI," in which algorithms can fulfil very specific tasks with high accuracy [32], the general conception of AI implementation is to use it as a tool to support clinicians in specific areas. However, it can be expected that with the development of AI in medicine, some tools might offer broader and more general applications. Different models of implementation of AI in clinical workflows have been conceived, in which the level of autonomy assigned to AI algorithms plays an important role. In this study, most patients reported that AI findings should be double-checked by a physician. In case of a disagreement between physicians and AI, the vast majority of patients (96.2% [153/159] for disagreement on diagnosis and 94.8% [146/154] for disagreement on treatment decisions) preferred the physician's to that of AI, concurrent with the results of a previous survey among Chinese cancer patients, in which 88.8% and 91.3% of participants reported they would follow the physician's suggestion regarding diagnosis and treatment [27].

The use of AI in medicine without adequate disclosure or explanation to patients can be hazardous [33,34]. Consequently, transparency and explicability are absolutely crucial for AI implementation [14]. Based on these findings on AI control and a significant trend towards lower acceptance rates of AI with increasing disease severity, we conclude that patients should be informed of which tasks involve AI algorithms and whether these applications are supervised by a physician. This is particularly relevant as most AI tools being developed and made available fall under this severe disease category, which explicitly comprised oncologic diseases in our questionnaire.

Our study has limitations that need to be acknowledged. The number of participants we included was rather small, which limits generalization of our results to other populations; for

example, a previous study suggested that Asian populations may anticipate a more disruptive development of medical AI with the potential replacement of health care professionals [35]. Apart from prior knowledge of AI, we did not observe other significant influences on its acceptance in the medical workflow, which might be attributed to the small sample. However, considering previous reports, we speculate that familiarity with AI technology is indeed the most important factor influencing patients' acceptance of it. Another limitation to consider is that the setting of handing out the questionnaire at registration for cross-sectional imaging, owing to organizational prerequisites, certainly introduced a selection bias towards participants with a history of more severe diseases warranting such radiological examinations. Surveying patients at primary care physician appointments might therefore yield divergent results.

In conclusion, patients had greater confidence in physicians than in AI in most clinical capabilities except for making treatment decisions based on the most recent scientific evidence, where they found AI advantageous. Patients strongly preferred physician-controlled application of AI. In order to safeguard patient interests, disclosure and control of AI application in medicine is crucial.

Acknowledgments

This study has been partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – LE 4401/1-1 to SL (project number 426969820), the Else Kröner-Fresenius Stifung (2018_EKMS.34 to NGH) and by the Koeln Fortune Program / Faculty of Medicine, University of Cologne (330/2020).

We thank Melanie Treutlein and Nedim Beste for their support with data collection.

Conflicts of Interest

None declared.

References

1. Girasa R. AI as a Disruptive Technology. In: Artif Intell as a Disruptive Technol. Cham, Switzerland: Palgrave Macmillan; 2020:3-21.
2. Graffy PM, Sandfort V, Summers RM, Pickhardt PJ. Automated Liver Fat Quantification at Nonenhanced Abdominal CT for Population-based Steatosis Assessment. Radiology 2019 Nov;293(2):334-342. [doi: 10.1148/radiol.2019190512] [Medline: 31526254]
3. Hinton G. Deep Learning-A Technology With the Potential to Transform Health Care. JAMA 2018 Sep 18;320(11):1101-1102. [doi: 10.1001/jama.2018.11100] [Medline: 30178065]
4. Wang C, Zhu X, Hong JC, Zheng D. Artificial Intelligence in Radiotherapy Treatment Planning: Present and Future. Technol Cancer Res Treat 2019 Jan 01;18:1533033819873922 [FREE Full text] [doi: 10.1177/1533033819873922] [Medline: 31495281]
5. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Feb 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
6. Chang HY, Jung CK, Woo JI, Lee S, Cho J, Kim SW, et al. Artificial Intelligence in Pathology. J Pathol Transl Med 2019 Jan;53(1):1-12 [FREE Full text] [doi: 10.4132/jptm.2018.12.16] [Medline: 30599506]
7. Alexander A, McGill M, Tarasova A, Ferreira C, Zurkiya D. Scanning the Future of Medical Imaging. J Am Coll Radiol 2019 Apr;16(4 Pt A):501-507 [FREE Full text] [doi: 10.1016/j.jacr.2018.09.050] [Medline: 30522844]
8. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018 Aug;18(8):500-510 [FREE Full text] [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
9. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, et al. Current Applications and Future Impact of Machine Learning in Radiology. Radiology 2018 Aug;288(2):318-328 [FREE Full text] [doi: 10.1148/radiol.2018171820] [Medline: 29944078]
10. Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P, et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison With 101 Radiologists. J Natl Cancer Inst 2019 Sep 01;111(9):916-922 [FREE Full text] [doi: 10.1093/jnci/djy222] [Medline: 30834436]
11. Ström P, Kartasalo K, Olsson H, Solorzano L, Delahunt B, Berney DM, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. Lancet Oncol 2020 Feb;21(2):222-232. [doi: 10.1016/S1470-2045(19)30738-7] [Medline: 31926806]
12. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature 2020 Jan;577(7788):89-94. [doi: 10.1038/s41586-019-1799-6] [Medline: 31894144]
13. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019 Jun;25(6):954-961. [doi: 10.1038/s41591-019-0447-x] [Medline: 31110349]

14. Geis JR, Brady AP, Wu CC, Spencer J, Ranschaert E, Jaremko JL, et al. Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement. J Am Coll Radiol 2019 Nov;16(11):1516-1521. [doi: 10.1016/j.jacr.2019.07.028] [Medline: 31585696]
15. Pinto Dos Santos D, Giese D, Brodehl S, Chon SH, Staab W, Kleinert R, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol 2019 Apr;29(4):1640-1646. [doi: 10.1007/s00330-018-5601-1] [Medline: 29980928]
16. European Society of Radiology (ESR). What the radiologist should know about artificial intelligence - an ESR white paper. Insights Imaging 2019 Apr 04;10(1):44 [FREE Full text] [doi: 10.1186/s13244-019-0738-2] [Medline: 30949865]
17. Arieno A, Chan A, Destounis SV. A Review of the Role of Augmented Intelligence in Breast Imaging: From Automated Breast Density Assessment to Risk Stratification. AJR Am J Roentgenol 2019 Feb;212(2):259-270. [doi: 10.2214/AJR.18.20391] [Medline: 30422711]
18. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient Perspectives on the Use of Artificial Intelligence for Skin Cancer Screening: A Qualitative Study. JAMA Dermatol 2020 May 01;156(5):501-512. [doi: 10.1001/jamadermatol.2019.5014] [Medline: 32159733]
19. Palmisciano P, Jamjoom AAB, Taylor D, Stoyanov D, Marcus HJ. Attitudes of Patients and Their Relatives Toward Artificial Intelligence in Neurosurgery. World Neurosurg 2020 Jun;138:e627-e633. [doi: 10.1016/j.wneu.2020.03.029] [Medline: 32179185]
20. Haan M, Ongena YP, Hommes S, Kwee TC, Yakar D. A Qualitative Study to Understand Patient Perspective on the Use of Artificial Intelligence in Radiology. J Am Coll Radiol 2019 Oct;16(10):1416-1419. [doi: 10.1016/j.jacr.2018.12.043] [Medline: 30878311]
21. Tran V, Riveros C, Ravaud P. Patients' views of wearable devices and AI in healthcare: findings from the ComPaRe e-cohort. NPJ Digit Med 2019;2:53 [FREE Full text] [doi: 10.1038/s41746-019-0132-y] [Medline: 31304399]
22. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
23. Allaire J. RStudio: integrated development environment for R. RStudio. 2012. URL: http://www.rstudio.com/ [accessed 2021-01-28]
24. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp 2018 Oct 24;2(1):35 [FREE Full text] [doi: 10.1186/s41747-018-0061-6] [Medline: 30353365]
25. Banja J. AI Hype and Radiology: A Plea for Realism and Accuracy. Radiology: Artificial Intelligence 2020 Jul 01;2(4):e190223. [doi: 10.1148/ryai.2020190223]
26. Fontelo P, Liu F. A review of recent publication trends from top publishing countries. Syst Rev 2018 Sep 27;7(1):147 [FREE Full text] [doi: 10.1186/s13643-018-0819-1] [Medline: 30261915]
27. Yang K, Zeng Z, Peng H, Jiang Y. Attitudes Of Chinese Cancer Patients Toward The Clinical Use Of Artificial Intelligence. Patient Prefer Adherence 2019;13:1867-1875 [FREE Full text] [doi: 10.2147/PPA.S225952] [Medline: 31802856]
28. Smith A. Public Attitudes Toward Computer Algorithms. Pew Research Center Internet & Technology. 2018. URL: https://www.pewresearch.org/internet/2018/11/16/public-attitudes-toward-computer-algorithms/ [accessed 2021-01-28]
29. Kuo W, H ne C, Mukherjee P, Malik J, Yuh EL. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. Proc Natl Acad Sci U S A 2019 Nov 05;116(45):22737-22745 [FREE Full text] [doi: 10.1073/pnas.1908021116] [Medline: 31636195]
30. Truong K. Aidoc gets FDA nod for AI pulmonary embolism screening tool. MedCity News. 2019. URL: https://medcitynews.com/2019/05/aidoc-gets-fda-nod-for-ai-pulmonary-embolism-screening-tool/ [accessed 2021-01-28]
31. Shefayim K. Zebra Medical Vision Receives FDA Approval for World's First AI Chest X-ray Triage Product. Zebra Medical Vision. 2019. URL: https://www.prnewswire.com/news-releases/zebra-medical-vision-receives-fda-approval-for-worlds-first-ai-chest-x-ray-triage-product-300848867.html [accessed 2021-01-28]
32. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
33. Cohen IG. Informed Consent and Medical Artificial Intelligence: What to Tell the Patient? SSRN Journal 2020;108:1425-1469. [doi: 10.2139/ssrn.3529576]
34. Robins R, Brodwin E. An invisible hand: Patients aren?t being told about the AI systems advising their care Internet. 2020. 2020. URL: https://www.statnews.com/2020/07/15/artificial-intelligence-patient-consent-hospitals/ [accessed 2021-01-27]
35. Gao S, He L, Chen Y, Li D, Lai K. Public Perception of Artificial Intelligence in Medical Care: Content Analysis of Social Media. J Med Internet Res 2020 Jul 13;22(7):e16649 [FREE Full text] [doi: 10.2196/16649] [Medline: 32673231]

Abbreviations

AI: artificial intelligence
CT: computed tomography

MRI: magnetic resonance imaging

Edited by G Eysenbach; submitted 10.09.20; peer-reviewed by F Stoehr, R Watson; comments to author 29.09.20; revised version received 01.11.20; accepted 30.11.20; published 17.02.21.

Please cite as:
Lennartz S, Dratsch T, Zopfs D, Persigehl T, Maintz D, Große Hokamp N, Pinto dos Santos D
Use and Control of Artificial Intelligence in Patients Across the Medical Workflow: Single-Center Questionnaire Study of Patient Perspectives
J Med Internet Res 2021;23(2):e24221
URL: http://www.jmir.org/2021/2/e24221/
doi:10.2196/24221
PMID:33595451

Original Paper

Automated Computer Vision Assessment of Hypomimia in Parkinson Disease: Proof-of-Principle Pilot Study

Avner Abrami[1], MSc; Steven Gunzler[2], MD; Camilla Kilbane[2], MD; Rachel Ostrand[1], PhD; Bryan Ho[3], MD; Guillermo Cecchi[1], PhD

[1]IBM Research – Computational Biology Center, Yorktown Heights, NY, United States

[2]Parkinson's and Movement Disorders Center, Neurological Institute, University Hospitals Cleveland Medical Center, Cleveland, OH, United States

[3]Department of Neurology, Tufts Medical Center, Boston, MA, United States

Corresponding Author:
Guillermo Cecchi, PhD
IBM Research – Computational Biology Center
1101 Kitchawan Rd
Yorktown Heights, NY, 10598
United States
Phone: 1 1 914 945 1815
Email: gcecchi@us.ibm.com

Abstract

Background: Facial expressions require the complex coordination of 43 different facial muscles. Parkinson disease (PD) affects facial musculature leading to "hypomimia" or "masked facies."

Objective: We aimed to determine whether modern computer vision techniques can be applied to detect masked facies and quantify drug states in PD.

Methods: We trained a convolutional neural network on images extracted from videos of 107 self-identified people with PD, along with 1595 videos of controls, in order to detect PD hypomimia cues. This trained model was applied to clinical interviews of 35 PD patients in their on and off drug motor states, and seven journalist interviews of the actor Alan Alda obtained before and after he was diagnosed with PD.

Results: The algorithm achieved a test set area under the receiver operating characteristic curve of 0.71 on 54 subjects to detect PD hypomimia, compared to a value of 0.75 for trained neurologists using the United Parkinson Disease Rating Scale-III Facial Expression score. Additionally, the model accuracy to classify the on and off drug states in the clinical samples was 63% (22/35), in contrast to an accuracy of 46% (16/35) when using clinical rater scores. Finally, each of Alan Alda's seven interviews were successfully classified as occurring before (versus after) his diagnosis, with 100% accuracy (7/7).

Conclusions: This proof-of-principle pilot study demonstrated that computer vision holds promise as a valuable tool for PD hypomimia and for monitoring a patient's motor state in an objective and noninvasive way, particularly given the increasing importance of telemedicine.

(J Med Internet Res 2021;23(2):e21037) doi:10.2196/21037

KEYWORDS

Parkinson disease; hypomimia; computer vision; telemedicine

Introduction

Facial expressions are an essential component of interpersonal communication [1]. They depend on our ability to voluntarily and involuntarily contract facial muscles [2], which are innervated by facial nerves. However, neurodegenerative diseases can cause cognitive disorders that affect expressivity [3] and cortical or peripheral nerve traumas [4], and can limit the production of facial expressions and emotion recognition [5,6]. This affects the ability to contract facial muscles by causing hemifacial spasms and can produce involuntary movements (ie, tics) and muscle weakness or stiffness.

Parkinson disease (PD) is a neurodegenerative disease that produces a gradual and generalized loss of motor functions, including the ability to contract facial muscles during spontaneous and voluntary emotional expressions [7], and voluntary nonemotional facial movements [8]. This reduced ability leads to a loss of facial expressiveness that generates a

signature "mask-like" appearance of the disease, which is also known as hypomimia. This loss of expressivity is often confounded with depression [2,9], a common symptom in patients with PD. However, even nondepressed PD patients show hypomimia, supporting the hypothesis of a motor control impairment in addition to the effects of depression [2,9]. A hypomimia rating is part of the Unified Parkinson Disease Rating Scale (UPDRS) [10], which is the gold standard clinical assessment tool. When assessing a patient for hypomimia, neurologists rate on a 5-point scale as follows: 0 for normal facial expression, 1 for minimal hypomimia, 2 for slight but abnormal diminution of facial expression, 3 for moderate hypomimia, and 4 for severe or complete loss of facial expression [10].

Disease progression does not seem to produce uniform facial masking across people. Studies of differential deficits in specific muscles [7] and sections of the face have documented asymmetric patterns [11] during posed smiling. However, previous work was based on constrained laboratory tasks where facial expressions were either evoked by sensory stimulations [12] or posed [7,11], limiting the applicability of the results to spontaneous natural expressivity.

In experimental settings, the quantification of masked facies in patients with PD has been traditionally performed with manual scoring [11,13,14]. A method capable of objectively characterizing variations in naturally occurring facial expressions that vary with disease progression would allow the patient state to be continuously evaluated outside of a clinical setting, opening up the possibility of remote or telemedicine-based monitoring. Video analysis has started to demonstrate success in objectively quantifying emotions in psychiatry [15,16] and neurology [17].

Computational methods based on known face components (eyes, mouth/lips, action units, skin, and shape) have been proposed [18-24]. For instance, eye-tracking algorithms [18] have been successful at quantifying the reduction in emotion recognition by patients with PD [19]. These methods involve the analysis of facial movements [22], specific visual features [23], patterns related to a specific emotion such as disgust [12], or facial landmarks [21]. Engineered facial features are nevertheless limited by image quality (distance to camera), pose (nonfrontal looking participant), or visual occlusions (eg, glasses, hands, and hats). These challenges [25] can be overcome by learning features directly from the raw image data using deep convolution networks that are known to be very effective at extracting emotions in healthy participants [26] and to outperform classic feature extraction methods.

Although it is well accepted that PD produces a generalized loss of the ability to produce facial expressions, it is unclear how this deficit evolves with disease progression and what are the effects of dopamine replacement therapy on masked facies [27]. In this work, we describe a methodology to characterize PD hypomimia using deep learning. This procedure can be performed remotely on videos, and thus, it provides a novel noninvasive digital tool for objective assessment of PD hypomimia and the changes associated with an *on-off* drug motor state. An automated video-based assessment tool like this one may be valuable for use in telemedicine [28], which has become increasingly utilized in PD especially following the onset of the COVID-19 pandemic. Such a resource would also allow for monitoring of a patient's motor state at home [29,30] between in-office neurologist or clinical trial visits [31].

Methods

Algorithm Development

The neural network model was trained using two data sets of faces, comprising people with PD and controls. The first was the YouTube Faces Database [32] (created by the Computer Science Department of Tel Aviv University), which contains 3425 videos of 1595 people (two-thirds of the subjects are male). The average length of the video clips is 7 seconds. This database constituted the control database for training the Visual Geometry Group neural network [33] in this study. The second training data set was created by performing a search on YouTube using the search terms "Parkinson's disease" and "interview." From that search, 107 videos of self-identified PD patients (68 males, middle-aged and older patients) were collected. This latter YouTube set was randomly partitioned into a 75% training set (80/107 videos, 50 males) and 25% test set (27/107 videos, 18 males). By design, this training data set incorporated common image quality challenges (such as varying lightning conditions, poses, and occasional presence of motion blur).

To preprocess the videos, faces were extracted from each frame of each video. Thereafter, each image was converted to grayscale, the intensity was normalized (mean=0.51, standard deviation=0.25), and the image was resized to a standardized 224 × 224 pixels. The neural network was trained using stochastic gradient descent.

After training, for each new video in the test set, the algorithm assigned each frame a score between 0 and 1, based on the degree of hypomimia that was detected by the algorithm in that frame. The scores of all of the frames of a video together formed a density distribution for that video (Figure 1), which demonstrated the proportion of frames that are assigned each likelihood of hypomimia. It is important to note that not all frames are indicative of the disease state, as a patient with PD may well have some frames where he/she does not exhibit hypomimia. Thus, the probability distribution for each video (and thus each subject) had a different shape proportional to the underlying hypomimia severity. We hypothesize that a PD video will have more frames with a higher hypomimia score than will a control video. Similarly, we hypothesize that a patient with PD in the *off* drug state will have more frames with a higher hypomimia score than that same subject in the *on* drug state.

In order to classify each video, we needed to characterize this density distribution for each video as a single number. To do so, we took the fifth quantile (Q) of that video's frame score density distribution (other quantiles can be used without loss of generality as discussed in the Results section). A video that exhibits low hypomimia should have a positively skewed distribution, as the bulk of the probability mass will be closer to 0, and therefore, will have a lower value of Q. In contrast, a video that exhibits high hypomimia should have a negatively

skewed distribution, with the bulk of the probability mass closer to 1, and therefore, will have a higher value of Q. Thus, the value of Q can be used to characterize how strongly hypomimia is detected in a given video, by representing how far along the 0 to 1 continuum is required to achieve 5% of the video's frames. Using this metric, we hypothesize that control videos will have a relatively lower Q (more frames concentrated toward 0) and PD videos will have a relatively higher Q (more frames concentrated toward 1; Figure 2).

Figure 1. The preprocessing pipeline for the input videos. Faces are extracted, greyscaled, and normalized. Then, each frame in the video is assigned a probabilistic classification assignment from 0 to 1 representing the degree of hypomimia. Thus, each video is represented by a probability distribution of frame scores. SGD: stochastic gradient descent.

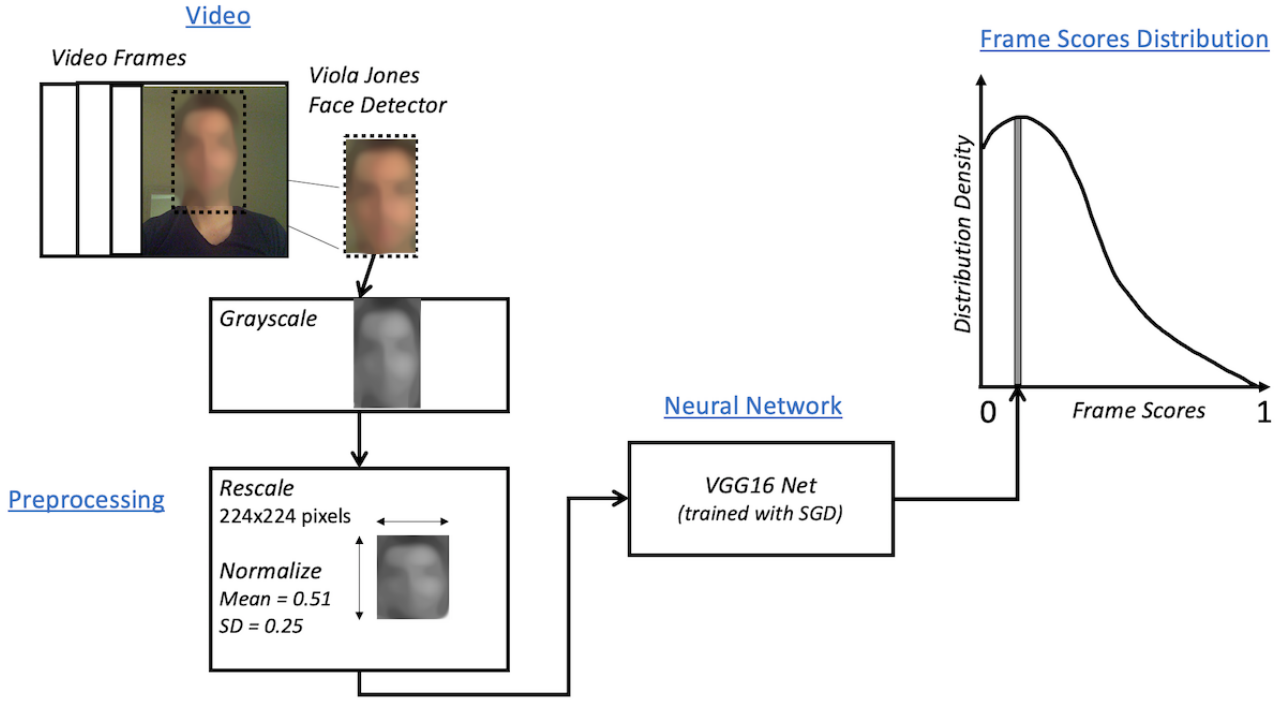
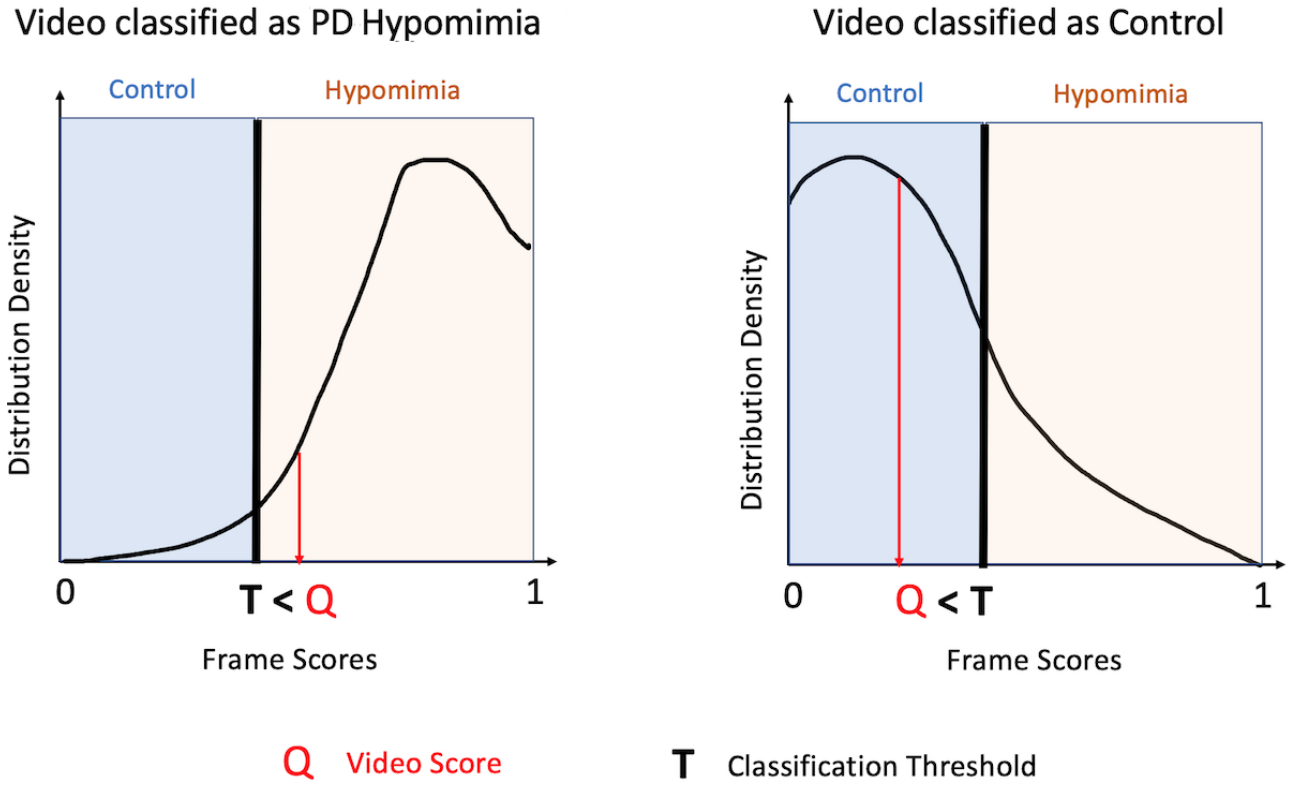


Figure 2. Video scoring. To classify a video, a probability distribution is created for all of a video's frames, and the fifth percentile of the distribution is defined as Q. A video that has a Q value above T (ie, closer to 1 or more evidence of hypomimia) is categorized as PD hypomimia; a video with a Q value below T (ie, closer to 0 or less evidence of hypomimia) is categorized as Control. PD: Parkinson disease.



Finally, a classification threshold T was selected. Any video that had a Q value lower than the threshold T was classified as not PD (ie, 0). Any video that had a Q value higher than the threshold T was classified as PD (ie, 1). T was selected such that it maximized classification accuracy in the testing set and was validated using the separate held-out validation set consisting of the Alan Alda videos.

Algorithm Testing

The difference in video scores between the PD and control groups was tested on a set of 54 videos (middle-aged and older patients, 37 males). Of these, half (n=27) featured people with self-identified PD, and the other half (n=27) featured people without PD (controls). The control videos were selected to include people who self-reported having other neurological or psychiatric disorders, with the following breakdown: 18 healthy people, four people with depression, one person with posttraumatic stress disorder, one person with traumatic brain injury, one person with bipolar disorder, one person with schizophrenia, and one person with chronic back pain. For the videos that were categorized as PD or other disorders, identification was performed based on the uploader's self-report (ie, the title of the video), not a clinical evaluation. However, many of the videos were created by disease associations, clinicians, academics, documentaries, or celebrities who publicly revealed their diagnoses, providing some degree of confidence of the reliability of the self-report.

Algorithm Validation

Hypomimia and Drug State

The Tufts Clinical data set consists of 35 participants (mean age 68 years, SD 8 years; 23 males and 12 females; mean total UPDRS-III score 25, SD 13) with idiopathic PD. The protocol was run at Tufts Medical Center in Boston, Massachusetts and was approved by the Tufts Health Sciences Campus Institutional Review Board (IRB #12371) (the complete study design [34] and related analyses conducted on the data set [35-37] have been reported previously). Patients were video recorded by means of a Microsoft Kinect camera (Microsoft Corp) at 30 frames per second.

Only 33 patients participated in a clinical interview in both their *on* and *off* drug states, with a mean of 3639 frames per video (approximately 2 minutes), which is similar to the length of the videos used in the training data set.

All 35 patients performed the UPDRS-III scripted tasks (including pronation-supination, finger tapping, and walking) and simulated activities of daily living [34] (including book carrying, bottle shaking, coat buttoning, cursive writing, and zipping) during their clinical visit, with a mean of 50,987 frames per video (approximately 28 minutes).

PD medication state (*on* or *off*) was self-reported by the participant at the start of each session. Medication dosage and timing was determined by the participant's typical daily dosage of levodopa (L-DOPA) therapy. Participants refrained from taking additional dosages in order to follow this experimental protocol. Participants were randomly assigned to an order condition (either completing the protocol in *on* first or *off* first). All participants arrived at the clinic in the *off* state. If they were assigned to the *off* first condition, they completed the experimental protocol when they arrived in the clinic (*off* state). Thereafter, they took their scheduled L-DOPA dose and waited until the medication's effects began. They were evaluated by the neurologist administering the UPDRS every 30 minutes until they self-reported being in the *on* medication state or 1.5 hours after the dose (whichever came earlier). Once this occurred, they completed the experimental protocol a second time (*on* state). In contrast, if the participant was assigned to the *on* first condition, they took their scheduled L-DOPA dose once they arrived at the clinic, waited until its effects began (as described above), and then completed the experimental protocol for the first time (*on* state). These participants then left the clinic and came back for a second scheduled session on a different day to perform the experimental protocol in the *off* state.

The UPDRS-III Facial Expression item, which rates the impairment of facial expressions, was used as the reference outcome variable in the present analyses. We characterized a strictly positive UPDRS-III Facial Expression score (ie, a rating greater than 0) as a positive PD classification by the examining neurologist. Additionally, we characterized a strictly positive difference of the UPDRS-III Facial Expression score between *off* and *on* (*off* minus *on*) as corresponding to a positive drug state classification by the neurologist.

Participants should have less dysfunction (and a lower UPDRS score) when they are in the *on* medication state than when they are in the *off* medication state. The present work investigates the effectiveness of the proposed computer vision algorithm to detect hypomimia in these patients, as well as quantify their medication state (ie, *on*/*off*) by detecting hypomimia. In that respect, if the algorithm is predictive of a patient's drug state, the model should predict a lower score for the *on* state as levodopa contributes to lowering PD symptoms by increasing the availability of dopamine to the brain. This hypothesis was tested by computing the change in score between the *off* and *on* medication states.

Longitudinal Severities of Masked Facies

The longitudinal data set consisted of seven videos of public appearances of Alan Alda from 1974 to 2019 (age 38-83 years), in which he was engaged in public speaking. Alan Alda is an actor, director, and screenwriter who was diagnosed with PD in 2014. This data set consists of four videos before diagnosis and three videos after diagnosis, and is used to evaluate the present algorithm's ability to quantify hypomimia. In this data set, a mean of 9642 frames per video (5.3 minutes) was extracted and analyzed by the algorithm. In these interviews, Mr Alda is recorded in diverse poses and lightning conditions, making the longitudinal data set qualitatively similar to the training data set and Tufts Clinical data set.

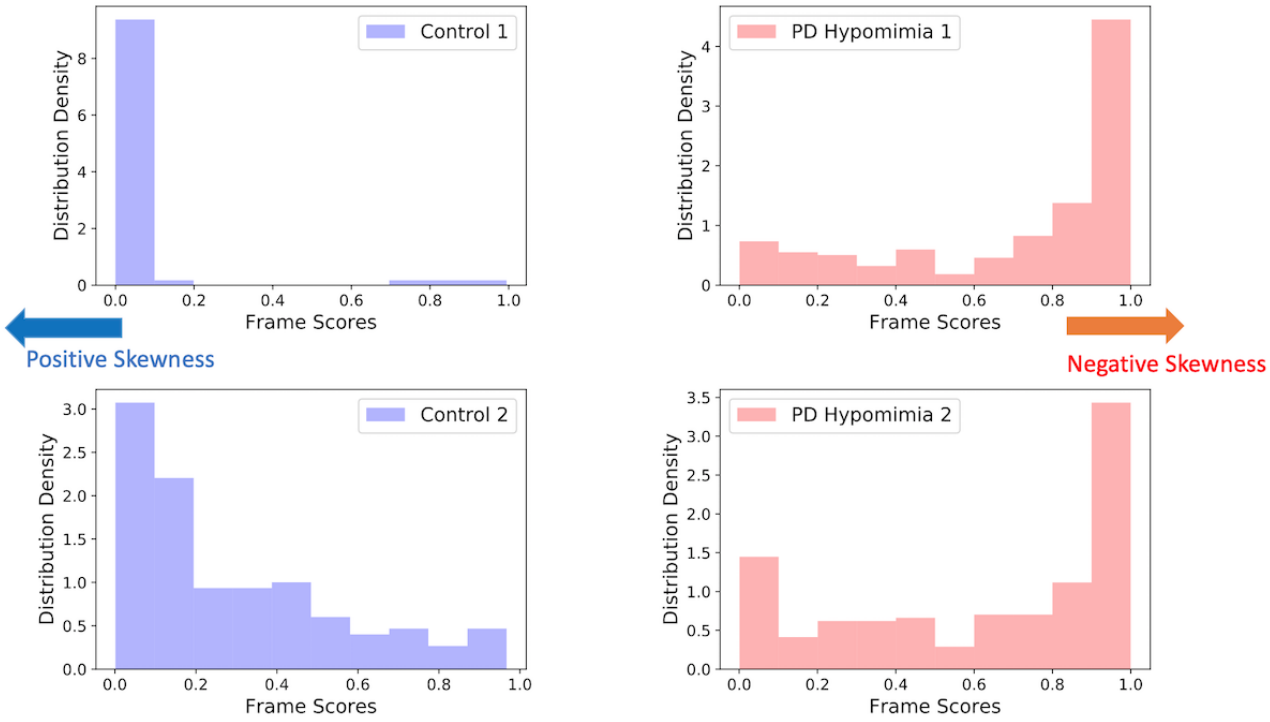
Results

Hypomimia Detection (Test Set)

As expected for the PD videos, a greater proportion of frames were classified as "PD hypomimia" than were for the control videos. The skewness of the PD subject video distributions was

significantly smaller than that of the control videos (one-tailed Mann-Whitney U=212, P=.004), demonstrating that there was more weight to the right side of the distribution (hypomimia scores closer to 1) for PD videos than for control videos (Figure 3).

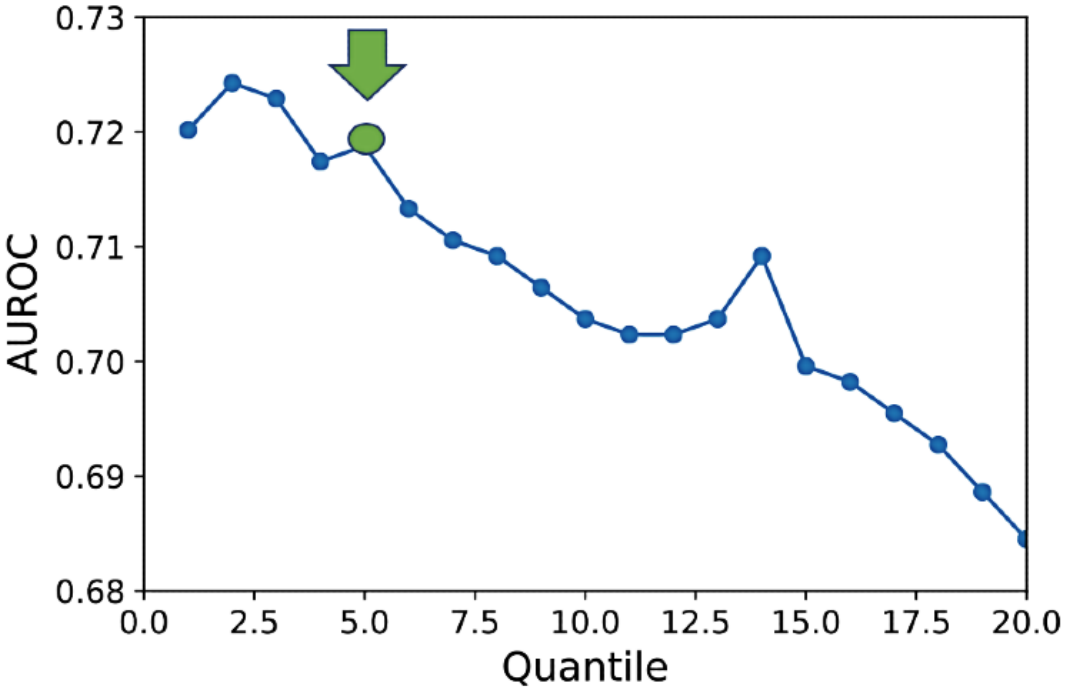
Figure 3. Example test set of PD and control distributions for four videos from the test data set. Two control videos (in blue) and two PD videos (in red). The distributions of the PD videos have higher weight on score values closer to 1 (negatively skewed) compared to the control videos, thereby demonstrating greater incidence of hypomimia. PD: Parkinson disease.



We experimentally quantify this difference in skewness by selecting the fifth quantile Q, which becomes the video score. The greater the incidence of hypomimia in the frames of a given video, the higher the quantile Q. Figure 4 shows that a wide range of quantiles would provide satisfying results on the test set (all quantiles below 15 achieve an area under the receiver operating characteristic curve [AUROC] >0.7). We chose the fifth quantile without loss of generality and applied this choice to the validation data sets only. A classification threshold (T=0.0003) was selected to maximize classification accuracy (70% accuracy or 38/54 videos correctly classified) in the test set. This threshold was determined on the basis of performance on these test set videos and then evaluated on the separate held-out validation data sets (Alan Alda and Tufts Clinical) to characterize hypomimia cues.

To provide a baseline accuracy measure, two professional neurologists rated each video in the test data set on the UPDRS-III Facial Expression score (score between 0 and 4). The neurologists performed the evaluation on the video, not an in-person clinical examination, and were told just to focus on the Facial Expression score and attempt to avoid being influenced by other cues present in the subject's behavior, to the extent possible. Using this scoring system, one neurologist's ratings produced an AUROC of 0.64 and the other neurologist's ratings produced an AUROC of 0.79. Averaging both neurologists' UPDRS-III Facial Expression scores produced an AUROC of 0.75. These scores were taken as an approximation of baseline classification accuracy that could be achieved using expert human raters. It is important to note, however, that this accuracy is an approximation and a true in-person clinical rating would incorporate substantially more information than just the UPDRS-III Facial Expression score.

Figure 4. Test set AUROC (PD vs Control) as a function of the chosen video distribution quantile to use as the "Q" threshold. A wide range of quantiles achieves AUROC > 0.7 (15th quantile and below). The fifth quantile, selected as our video score threshold, is shown in green. AUROC: area under the receiver operating characteristic curve; PD: Parkinson disease.



Hypomimia Changes Associated With the Drug State in the Tufts Data Set (Validation Set)

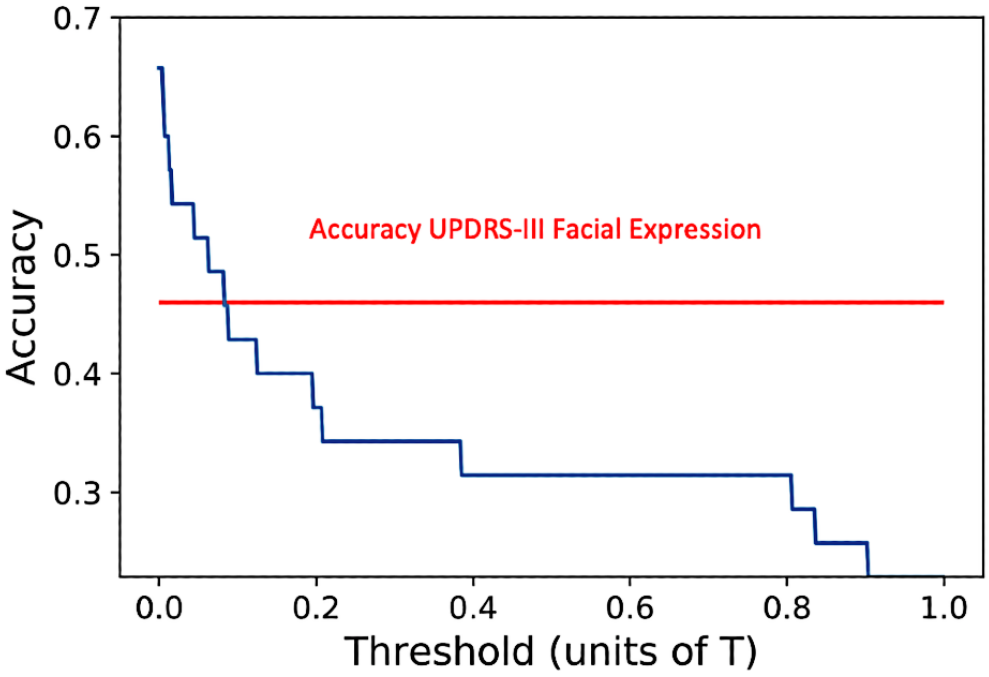
Finally, we assessed the performance of our algorithm on the held-out validation sets, after being trained on the training set and accuracy maximized on the testing set. The first validation set was the Tufts Clinical data set used to assess hypomimia changes associated with the drug state. For each patient, we extracted and analyzed all frames in the video. Our goal was to quantify PD hypomimia for each patient's visit and see if *on* and *off* motor states had an impact on hypomimia as captured by the neurologist's UPDRS-III Facial Expression score and our algorithm.

To test if the algorithm was able to correctly classify PD hypomimia, we computed the score of each clinical interview video to see if it exceeded the decision threshold T. If the score was above the threshold, the video was categorized as PD. The algorithm detected PD in 76% (25/33) of the *off* drug sessions (in comparison, the neurologist gave a UPDRS-III Facial Expression score higher than 0 for 88% [29/33] of these sessions) and in 67% (22/33) of the *on* drug sessions (in comparison, it was 70% [23/33] for the neurologist). This reduction in facial masking detection between the *off* and *on* drug sessions for both the algorithm and the neurologist can be attributed to drug efficacy in reducing PD symptoms.

To quantitatively evaluate the difference between the *off* and *on* states, we computed the difference between the *off* and *on* video scores for each participant. The *off* score generated by the algorithm was strictly greater than the *on* score in 63% (22/35) of the participants during the clinical visit. In comparison, the neurologist ratings of the UPDRS-III Facial Expression score were strictly greater in the *off* state than in the *on* state for only 46% (16/35) of participants. However, it is worth noting that the UPDRS-III Facial Expression score is integer based and does not allow clinicians to assess changes in facial expression that are more granular than these integer ratings. The clinical interview accuracies of the UPDRS-III Facial Expression score and algorithm were 45% (15/33) and 55% (18/33), respectively.

To quantify the sensitivity of our analysis, we provided a plot highlighting the differences in detection as given by different thresholds (the x-axis is scaled by T). Thresholds to separate the *on* state from the *off* state effectively were smaller than T, which appeared reasonable as we expected the difference between the *on* and *off* states to be more subtle than the difference between the PD and control groups. More precisely, a threshold of 0.1 T provided an accuracy comparable to that of the neurologist (46% accuracy), and a threshold of 0.01 T led to 60% accuracy (Figure 5).

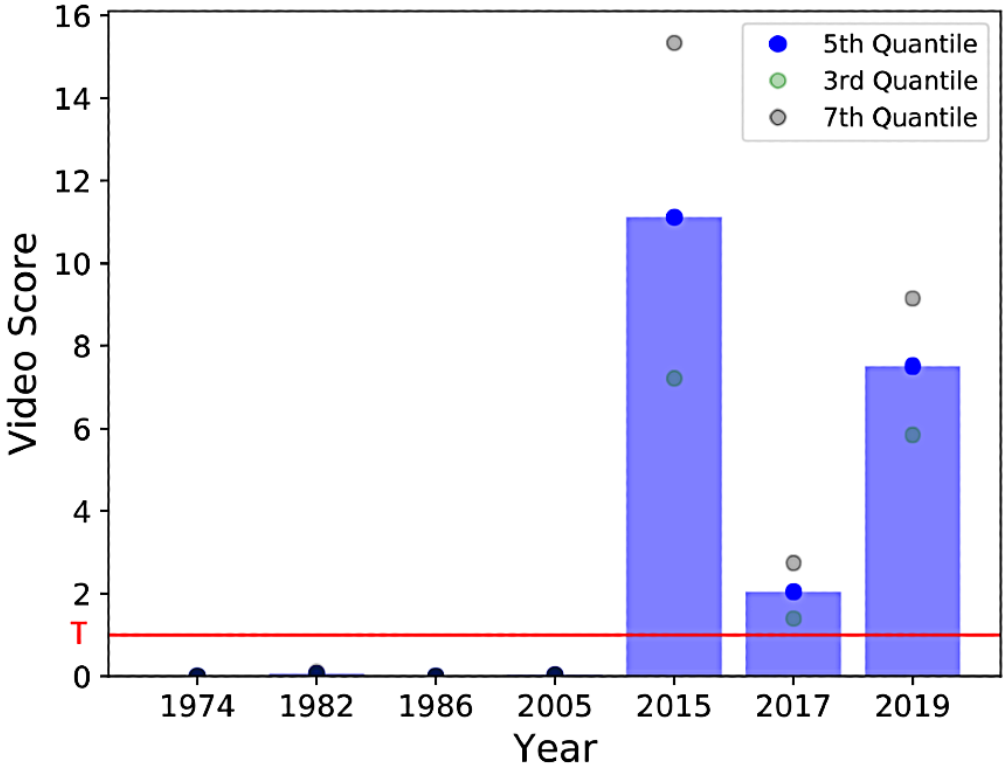
Figure 5. On-off sensitivity of the Tufts data set. On-off classification for a clinical visit is displayed as a function of threshold. On-off differences are far more subtle than PD versus Control differences (on which the model was originally trained). The red line shows the percentage of participants for which the neurologists rated the UPDRS-III Facial Expression score higher in the off state than in the on state. PD: Parkinson disease; UPDRS: United Parkinson Disease Rating Scale.



Longitudinal Severities of Masked Facies

We sought to retrospectively validate the algorithm's ability to characterize PD symptomology in an individual longitudinally. The algorithm was applied to seven interview videos featuring Alan Alda (officially diagnosed with PD in 2014) from 1974 to 2019. There was an increase in the algorithm's PD classification before PD diagnosis to after diagnosis. Indeed, all videos before PD diagnosis were below the optimal threshold T for positive classification, and all videos after diagnosis were well above the threshold, highlighting the fact that the algorithm was able to capture hypomimia cues. Finally, we included a confidence interval (as given by the third and seventh video distribution quantiles) associated with the video scores (Figure 6).

Figure 6. Validation with Alan Alda interviews. All videos after Mr Alda's PD diagnosis are above the threshold T, whereas videos before his PD diagnosis are below T (horizontal red line), indicating that the algorithm is sensitive to PD hypomimia symptomatology. Dots show the confidence interval (third quantile and seventh quantile of the video density distribution). PD: Parkinson disease.



Discussion

Principal Results

In this proof-of-principle pilot study, we used deep learning to detect PD hypomimia from videos of people with and without PD. Our method was also able to detect the effect of dopamine replacement medication in participants during their clinical visit and to analyze the progression of symptoms in the actor Alan Alda before and after his diagnosis of PD.

Comparison With Prior Work and an Alternative Approach

A well-established method to identify facial expressions was proposed by Ekman and Friesen [38], which describes visual facial movements related to the muscles involved in the production of emotions. Known as the Facial Action Coding System (FACS), this method uses localized image information and has been previously applied to study parkinsonism [12,39]. However, the studies did not provide information on medication state or longitudinal changes of facial expressions in PD patients.

An alternative approach to investigate the progression of PD as a function of the ability to move specific facial muscles is to use electromyography [7]. This approach may be considered less prone to artifacts of head movement, complexion, and facial bone structure, but the use of electromyography at a participant's home is technically challenging and impractical.

Limitations

There are noteworthy limitations to our work. The training data set is limited, as in particular, it did not include people in the full range of relevant ages affected by PD (early onset PD patients were not represented), which constrains generalizability. The effect of dopaminergic medication was not taken into consideration when training the model, as all videos in the training data sets were classified as either PD or control, with no consideration of *on* versus *off* medication state. There is uncertainty linked to the video labels in the training data set, as it relied on the uploader's self-report (ie, the title of the video), not a clinical evaluation. Consequently, it is important to validate the present algorithm with clinically verified participants with and without PD, as was performed in this work using clinically validated participants in *on* and *off* drug states. Moreover, while gender and, to a lesser extent, age can be ascertained with some degree of certainty for subjects in the training data set, additional demographic information is highly limited, making generalization to larger test sets potentially susceptible to demographic or other biases in the training sets. This observation, coupled with the fact that the longitudinal study is limited as it was applied on only seven videos of one person, implies that more extensively curated training data sets as well as larger testing data sets will be required to validate the robustness of our method.

Conclusions

Our algorithm may serve as a nonclinical marker for PD hypomimia and *on* and *off* motor states. Unlike the lengthy physical examination techniques required for the clinical assessment of PD, which require in-person or video examinations that must be rated by a trained clinician, the present automated technique is capable of rating videos of a patient's face. This technique has the potential to improve the ability to continuously monitor the *on* and *off* states, even in the

patient's home. This can thereby serve as a serial data point for use in at-home monitoring for PD or at-home assessment in a PD clinical research study, with little patient burden and minimal technological requirements.

With a shift toward a greater role of telemedicine, an automated assessment of hypomimia could serve as a screening tool for parkinsonism and as a nonobtrusive objective score to assess *on* and *off* states. Further study will be needed to assess the value of this automated assessment in various clinical settings. The proposed model was tested on PD hypomimia, but in theory, it could be applied to other neurological conditions that produce other face signatures.

Acknowledgments

The authors acknowledge the help provided by Dr John Gunstad in the initial phase of the project (encouraging our efforts and connecting us with clinical specialists in Parkinson disease) and the feedback provided by Dr Ajay Royyuru over several iterations of the results. The authors would also like to thank Alan Alda for allowing us to display the outcome of our algorithm on some of his publicly available videos.

Conflicts of Interest

None declared.

References

1. Frith C. Role of facial expressions in social interactions. Philos Trans R Soc Lond B Biol Sci 2009 Dec 12;364(1535):3453-3458 [FREE Full text] [doi: 10.1098/rstb.2009.0142] [Medline: 19884140]
2. Rinn WE. The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions. Psychol Bull 1984 Jan;95(1):52-77. [Medline: 6242437]
3. Bersani G, Polli E, Valeriani G, Zullo D, Melcore C, Capra E, et al. Facial expression in patients with bipolar disorder and schizophrenia in response to emotional stimuli: a partially shared cognitive and social deficit of the two disorders. Neuropsychiatr Dis Treat 2013;9:1137-1144 [FREE Full text] [doi: 10.2147/NDT.S46525] [Medline: 23966784]
4. Gordin E, Lee TS, Ducic Y, Arnaoutakis D. Facial nerve trauma: evaluation and considerations in management. Craniomaxillofac Trauma Reconstr 2015 Mar;8(1):1-13 [FREE Full text] [doi: 10.1055/s-0034-1372522] [Medline: 25709748]
5. Pick S, Goldstein LH, Perez DL, Nicholson TR. Emotional processing in functional neurological disorder: a review, biopsychosocial model and research agenda. J Neurol Neurosurg Psychiatry 2019 Jun;90(6):704-711 [FREE Full text] [doi: 10.1136/jnnp-2018-319201] [Medline: 30455406]
6. Shiota MN, Simpson ML, Kirsch HE, Levenson RW. Emotion recognition in objects in patients with neurological disease. Neuropsychology 2019 Nov;33(8):1163-1173 [FREE Full text] [doi: 10.1037/neu0000587] [Medline: 31478721]
7. Livingstone SR, Vezer E, McGarry LM, Lang AE, Russo FA. Deficits in the Mimicry of Facial Expressions in Parkinson's Disease. Front Psychol 2016;7:780 [FREE Full text] [doi: 10.3389/fpsyg.2016.00780] [Medline: 27375505]
8. Marneweck M, Hammond G. Voluntary control of facial musculature in Parkinson's disease. J Neurol Sci 2014 Dec 15;347(1-2):332-336. [doi: 10.1016/j.jns.2014.11.003] [Medline: 25467144]
9. Vogel HP. Symptoms of depression in Parkinson's disease. Pharmacopsychiatria 1982 Nov;15(6):192-196. [doi: 10.1055/s-2007-1019537] [Medline: 7156182]
10. Fahn S, Elton R. Unified Parkinson's Disease Rating Scale. In: Fahn S, Marsden CD, Goldstein M, Calne DB, editors. Recent Developments in Parkinson's Disease. Florham Park, NJ: Macmillan Healthcare Information; 1987:153-163.
11. Ricciardi L, Visco-Comandini F, Erro R, Morgante F, Volpe D, Kilner J, et al. Emotional facedness in Parkinson's disease. J Neural Transm (Vienna) 2018 Dec;125(12):1819-1827. [doi: 10.1007/s00702-018-1945-6] [Medline: 30343335]
12. Wu P, Gonzalez I, Patsis G, Jiang D, Sahli H, Kerckhofs E, et al. Objectifying facial expressivity assessment of Parkinson's patients: preliminary study. Comput Math Methods Med 2014;2014:427826 [FREE Full text] [doi: 10.1155/2014/427826] [Medline: 25478003]
13. Marsili L, Agostino R, Bologna M, Belvisi D, Palma A, Fabbrini G, et al. Bradykinesia of posed smiling and voluntary movement of the lower face in Parkinson's disease. Parkinsonism Relat Disord 2014 Apr;20(4):370-375. [doi: 10.1016/j.parkreldis.2014.01.013] [Medline: 24508573]
14. Gunnery SD, Naumova EN, Saint-Hilaire M, Tickle-Degnen L. Mapping spontaneous facial expression in people with Parkinson's disease: A multiple case study design. Cogent Psychol 2017 Sep 25;4(1):1376425 [FREE Full text] [doi: 10.1080/23311908.2017.1376425] [Medline: 29607351]
15. Fei Z, Yang E, Li D, Butler S, Ijomah W, Li S, et al. Deep convolution network based emotion analysis towards mental health care. Neurocomputing 2020 May;388:212-227 [FREE Full text] [doi: 10.1016/j.neucom.2020.01.034]
16. Leo M, Carcagnì P, Distante C, Spagnolo P, Mazzeo PL, Rosato AC, et al. Computational Assessment of Facial Expression Production in ASD Children. Sensors (Basel) 2018 Nov 16;18(11) [FREE Full text] [doi: 10.3390/s18113993] [Medline: 30453518]

17. Yolcu G, Oztel I, Kazan S, Oz C, Palaniappan K, Lever TE, et al. Facial expression recognition for monitoring neurological disorders based on convolutional neural network. Multimed Tools Appl 2019 Jul 23;78(22):31581-31603 [FREE Full text] [doi: 10.1007/s11042-019-07959-6]
18. Bek J, Poliakoff E, Lander K. Measuring emotion recognition by people with Parkinson's disease using eye-tracking with dynamic facial expressions. J Neurosci Methods 2020 Feb 01;331:108524. [doi: 10.1016/j.jneumeth.2019.108524] [Medline: 31747554]
19. Chen H, Wu D, He J, Li S, Su W. Emotion Recognition in Patients With Parkinson Disease. Cogn Behav Neurol 2019 Dec;32(4):247-255. [doi: 10.1097/WNN.0000000000000209] [Medline: 31800485]
20. Katsikitis M, Pilowsky I. A study of facial expression in Parkinson's disease using a novel microcomputer-based method. J Neurol Neurosurg Psychiatry 1988 Mar 01;51(3):362-366 [FREE Full text] [doi: 10.1136/jnnp.51.3.362] [Medline: 3361329]
21. Bandini A, Orlandi S, Escalante HJ, Giovannelli F, Cincotta M, Reyes-Garcia CA, et al. Analysis of facial expressions in parkinson's disease through video-based automatic methods. J Neurosci Methods 2017 Apr 01;281:7-20. [doi: 10.1016/j.jneumeth.2017.02.006] [Medline: 28223023]
22. Bowers D, Miller K, Bosch W, Gokcay D, Pedraza O, Springer U, et al. Faces of emotion in Parkinsons disease: micro-expressivity and bradykinesia during voluntary facial expressions. J Int Neuropsychol Soc 2006 Nov;12(6):765-773. [doi: 10.1017/S135561770606111X] [Medline: 17064440]
23. Gonzalez I, Sahli H, Enescu V, Verhelst W. Context-Independent Facial Action Unit Recognition Using Shape and Gabor Phase Information. In: D'Mello S, Graesser A, Schuller B, Martin JC, editors. Affective Computing and Intelligent Interaction. ACII 2011. Lecture Notes in Computer Science, vol 6974. Berlin, Heidelberg: Springer; 2011:548-557.
24. Leo M, Carcagnì P, Mazzeo P, Spagnolo P, Cazzato D, Distante C. Analysis of Facial Information for Healthcare Applications: A Survey on Computer Vision-Based Approaches. Information 2020 Feb 26;11(3):128 [FREE Full text] [doi: 10.3390/info11030128]
25. Kanade T, Cohn J, Tian Y. Comprehensive Database for Facial Expression Analysis. In: Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580). 2000 Presented at: Fourth IEEE International Conference on Automatic Face and Gesture Recognition; March 28-30, 2000; Grenoble, France. [doi: 10.1109/AFGR.2000.840611]
26. Leo M, Carcagnì P, Distante C, Mazzeo PL, Spagnolo P, Levante A, et al. Computational Analysis of Deep Visual Data for Quantifying Facial Expression Production. Applied Sciences 2019 Oct 25;9(21):4542. [doi: 10.3390/app9214542]
27. Bologna M, Fabbrini G, Marsili L, Defazio G, Thompson PD, Berardelli A. Facial bradykinesia. J Neurol Neurosurg Psychiatry 2013 Jun;84(6):681-685. [doi: 10.1136/jnnp-2012-303993] [Medline: 23236012]
28. Tan JR, Coulson S, Keep M. Face-to-Face Versus Video Assessment of Facial Paralysis: Implications for Telemedicine. J Med Internet Res 2019 Apr 12;21(4):e11109 [FREE Full text] [doi: 10.2196/11109] [Medline: 30977734]
29. Agnisarman S, Narasimha S, Chalil Madathil K, Welch B, Brinda F, Ashok A, et al. Toward a More Usable Home-Based Video Telemedicine System: A Heuristic Evaluation of the Clinician User Interfaces of Home-Based Video Telemedicine Systems. JMIR Hum Factors 2017 Apr 24;4(2):e11 [FREE Full text] [doi: 10.2196/humanfactors.7293] [Medline: 28438724]
30. Marceglia S, Rossi E, Rosa M, Cogiamanian F, Rossi L, Bertolasi L, et al. Web-based telemonitoring and delivery of caregiver support for patients with Parkinson disease after deep brain stimulation: protocol. JMIR Res Protoc 2015 Mar 06;4(1):e30 [FREE Full text] [doi: 10.2196/resprot.4044] [Medline: 25803512]
31. Pan D, Dhall R, Lieberman A, Petitti DB. A mobile cloud-based Parkinson's disease assessment system for home-based monitoring. JMIR Mhealth Uhealth 2015 Mar 26;3(1):e29 [FREE Full text] [doi: 10.2196/mhealth.3956] [Medline: 25830687]
32. Wolf L, Hassner T, Maoz I. Face recognition in unconstrained videos with matched background similarity. In: CVPR 2011. 2011 Presented at: Conference on Computer Vision and Pattern Recognition (CVPR); June 20-25, 2011; Providence, RI, USA. [doi: 10.1109/cvpr.2011.5995566]
33. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv. URL: https://arxiv.org/abs/1409.1556 [accessed 2021-02-02]
34. Erb K, Daneault J, Amato S, Bergethon P, Demanuele C, Kangarloo T, et al. The BlueSky Project: monitoring motor and non-motor characteristics of people with Parkinson's disease in the laboratory, a simulated apartment, and home and community settings. Mov Disord. 2018. URL: https://tinyurl.com/psw6wb9f [accessed 2021-02-02]
35. Erb MK, Karlin DR, Ho BK, Thomas KC, Parisi F, Vergara-Diaz GP, et al. mHealth and wearable technology should replace motor diaries to track motor fluctuations in Parkinson's disease. NPJ Digit Med 2020 Jan 17;3(1):6 [FREE Full text] [doi: 10.1038/s41746-019-0214-x] [Medline: 33452390]
36. Abrami A, Heisig S, Ramos V, Thomas KC, Ho BK, Caggiano V. Using an unbiased symbolic movement representation to characterize Parkinson's disease states. Sci Rep 2020 Apr 30;10(1):7377 [FREE Full text] [doi: 10.1038/s41598-020-64181-3] [Medline: 32355166]
37. Mahadevan N, Demanuele C, Zhang H, Volfson D, Ho B, Erb MK, et al. Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device. NPJ Digit Med 2020 Jan 15;3(1):5 [FREE Full text] [doi: 10.1038/s41746-019-0217-7] [Medline: 33441944]

38. Ekman P, Friesen W. Facial Action Coding System: A Technique for the Measurement of Facial Movement. 1978. URL: https://www.semanticscholar.org/paper/Facial-action-coding-system%3A-a-technique-for-the-of-Ekman-Friesen/1566cf20e2ba91ca8857c30083419bf7c127094b [accessed 2021-02-02]
39. Vinokurov N, Arkadir D, Linetsky E, Bergman H, Weinshall D. Quantifying Hypomimia in Parkinson Patients Using a Depth Camera. In: Serino S, Matic A, Giakoumis D, Lopez G, Cipresso P, editors. Pervasive Computing Paradigms for Mental Health. MindCare 2015. Communications in Computer and Information Science, vol 604. Cham: Springer; 2016:63-71.

Abbreviations

AUROC: area under the receiver operating characteristic curve
L-DOPA: levodopa
PD: Parkinson disease
UPDRS: United Parkinson Disease Rating Scale

Edited by G Eysenbach; submitted 04.06.20; peer-reviewed by R Dorsey, S Guness, M Leo; comments to author 24.06.20; revised version received 30.07.20; accepted 18.12.20; published 22.02.21.

Please cite as:
Abrami A, Gunzler S, Kilbane C, Ostrand R, Ho B, Cecchi G
Automated Computer Vision Assessment of Hypomimia in Parkinson Disease: Proof-of-Principle Pilot Study
J Med Internet Res 2021;23(2):e21037
URL: https://www.jmir.org/2021/2/e21037
doi:10.2196/21037
PMID:33616535

Original Paper

A Risk Prediction Model Based on Machine Learning for Cognitive Impairment Among Chinese Community-Dwelling Elderly People With Normal Cognition: Development and Validation Study

Mingyue Hu[1], MSc; Xinhui Shu[2], MSc; Gang Yu[3], PhD; Xinyin Wu[4], PhD; Maritta Välimäki[1,5*], PhD; Hui Feng[1,6,7*], PhD

[1]Xiangya Nursing School, Central South University, Changsha, China

[2]Henan Cancer Hospital Province, Zhengzhou University, Zhengzhou, China

[3]Department of Biomedical Engineering, School of Basic Medical Science, Central South University, Changsha, China

[4]Xiangya School of Public Health, Central South University, Changsha, China

[5]Department of Nursing Science, University of Turku, Helsinki, Finland

[6]Oceanwide Health Management Institute, Central South University, Changsha, China

[7]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Hui Feng, PhD
Xiangya Nursing School
Central South University
Yuelu District
172 Tongzipo Road
Changsha
China
Phone: 86 15173121969
Email: feng.hui@csu.edu.cn

Abstract

Background: Identifying cognitive impairment early enough could support timely intervention that may hinder or delay the trajectory of cognitive impairment, thus increasing the chances for successful cognitive aging.

Objective: We aimed to build a prediction model based on machine learning for cognitive impairment among Chinese community-dwelling elderly people with normal cognition.

Methods: A prospective cohort of 6718 older people from the Chinese Longitudinal Healthy Longevity Survey (CLHLS) register, followed between 2008 and 2011, was used to develop and validate the prediction model. Participants were included if they were aged 60 years or above, were community-dwelling elderly people, and had a cognitive Mini-Mental State Examination (MMSE) score ≥18. They were excluded if they were diagnosed with a severe disease (eg, cancer and dementia) or were living in institutions. Cognitive impairment was identified using the Chinese version of the MMSE. Several machine learning algorithms (random forest, XGBoost, naïve Bayes, and logistic regression) were used to assess the 3-year risk of developing cognitive impairment. Optimal cutoffs and adjusted parameters were explored in validation data, and the model was further evaluated in test data. A nomogram was established to vividly present the prediction model.

Results: The mean age of the participants was 80.4 years (SD 10.3 years), and 50.85% (3416/6718) were female. During a 3-year follow-up, 991 (14.8%) participants were identified with cognitive impairment. Among 45 features, the following four features were finally selected to develop the model: age, instrumental activities of daily living, marital status, and baseline cognitive function. The concordance index of the model constructed by logistic regression was 0.814 (95% CI 0.781-0.846). Older people with normal cognitive functioning having a nomogram score of less than 170 were considered to have a low 3-year risk of cognitive impairment, and those with a score of 170 or greater were considered to have a high 3-year risk of cognitive impairment.

Conclusions: This simple and feasible cognitive impairment prediction model could identify community-dwelling elderly people at the greatest 3-year risk for cognitive impairment, which could help community nurses in the early identification of dementia.

(J Med Internet Res 2021;23(2):e20298) doi:10.2196/20298

KEYWORDS

prediction model; cognitive impairment; machine learning; nomogram

Introduction

Dementia constitutes a major health care burden nationally and worldwide [1]. Approximately every 3 seconds, a person somewhere in the world is diagnosed with dementia, and the current annual cost of dementia is estimated to be US $1 trillion, which is set to double by 2030 [2]. China has the largest population of patients with dementia in the world (9.5 million) followed by the United States (4.2 million) [1]. Evidence suggests that delaying the onset of dementia by 1 year is likely to reduce its prevalence by 11% by 2050, while delaying it by 5 years could halve the number of people living with dementia by 2050 [3]. Given that dementia is incurable, it is of high importance to detect cognitive impairment in its early stages [4].

Good evidence already exists that specific risk factors can contribute to increased dementia risk at different life stages. The risk factors are education in early life, hypertension and obesity in midlife, and smoking and depression in later life [5]. Prediction models concerning risk factors for cognitive impairment have already been published. However, the variables included in the models vary, and they mostly focus on laboratory markers only [6-8]. A systematic review by Hou et al included 61 studies of the prediction models of dementia. They found that age, sex, education, cognition assessment scales, BMI, alcohol intake, and genetic variables were the most common predictors included in the models [8]. Questionnaire-based data have also been used to explore the clinical variables with promising predictive values in the transition to cognitive impairment (demographic characteristics and neuropsychiatric symptoms). Other studies have used data based on medical imaging (brain atrophy), genes (apolipoprotein Eε4), or biomarkers (amyloid-β, tau, etc) [5,9,10]. One study used the C-Pittsburgh compound B (C-PiB) medial temporal standard uptake value ratio with the Mini-Mental State Examination (MMSE) for the prediction of a person going from mild cognitive impairment to dementia, and the area under the curve was 0.92 [6]. Kivipelto et al used big data to develop a prediction model of the risk of late-life dementia in middle-aged people, and the model included age, education, hypertension, hypercholesterolemia, and obesity as variables, with an average area under the curve of 0.77 [7]. However, these prediction models are complex, less accurate, and difficult to implement in practice for nursing staff who are working with elderly patients. Therefore, especially for use in community environments, simpler, more accurate, and feasible models are needed [8].

Machine learning has recently been used to produce a prediction model for practice. Machine learning can help in modeling information based on causal and/or statistical data, potentially revealing hidden dependencies between factors and diseases in a big data environment [11]. Published studies show how machine learning algorithms, such as naïve Bayes (NB), AdaBoost, and random forest (RF), have been used to predict or detect cognitive impairment [12-15].

We systematically searched PubMed (["cognitive impairment" OR "cognitive decline" OR "dementia" OR "alzheimer*"] AND ["machine learning" OR "data mining" OR "big data"] AND "prediction") and found four studies in which machine learning was used to identify risk factors for dementia among people with normal cognition at baseline. One study [16] used unsupervised machine learning to develop a dementia prediction model that could identify people at a high risk of developing dementia. Another study [13] used the medical records of 93,120 patients to develop a model for exploring undetected dementia using a machine learning approach (with an area under the curve of 0.74). One study [17] developed a model for predicting the risk of developing dementia within the next 2 years among older people (aged 85 years or above) without dementia (with an area under the curve of 0.73). The study showed that the predictors differed between the youngest and oldest individuals in the population. Further, another study used supervised machine learning to develop a dementia prediction model (area under the curve values of 0.75 and 0.79) and found that the Disease State Index is useful for identifying individuals who are most at risk [18].

However, a variety of difficulties have been identified in implementing the results of machine learning in clinical practice, as the data have been collected at one time point only, meaning that the causality of the data can be questioned [13]. Some prediction models have been too complicated, and there have been problems with accuracy in the prediction [16-18]. In addition, although the results seem to be acceptable from a statistical point of view, understanding the interpretation of the unsupervised machine learning result and its implementation into practice is demanding [16]. There is still room for the improvement of prediction models for forecasting risks for dementia. In addition, more studies are needed to develop and translate the results into clinical practice, especially for community environments [19]. We therefore aimed to develop an algorithm to be used in a prediction model to identify risk factors for cognitive impairment among Chinese community-dwelling elderly people with normal cognition. The study results are important, as an approach to stratify the individual risks for cognitive impairment is needed in community settings for both national and international purposes [20].

Methods

Design and Participants

This study strictly followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement [21]. In this machine learning approach, the national prospective longitudinal results of the Chinese Longitudinal Healthy Longevity Survey

(CLHLS) were used [22]. The CLHLS is one of the largest national longitudinal studies for investigating the health of older Chinese adults. Launched in 1998, the CLHLS implemented follow-up surveys in 2000, 2002, 2005, 2008-09, 2011-12, and 2014. A total of 22 Chinese provinces were randomly recruited, and the sampling frame covered about 85% of the total population of China. The survey results in the national database are freely accessible and available online [23]. The 2008-09 survey included a total of 16,954 participants.

We included 11,788 participants from the 2008-09 wave, and 6718 participants were eligible for model development and internal validation. Participants were included if they were (1) aged 60 years or above; (2) community-dwelling elderly people; and (3) normally cognitive (MMSE score ≥18). They were excluded if they (1) were diagnosed with a severe disease (eg, cancer and dementia) or (2) lived in an institution. A detailed flow chart of participant selection is shown in Multimedia Appendix 1. Among the remaining participants, in the 2011-12 wave, a total of 1913 participants were lost in the follow-up and 2879 died. Those who were excluded from analyses owing to nonparticipation or death were on average older (P<.001) and had lower physical function scores (P<.001) and lower baseline cognition scores (P<.001). The two groups were not significantly different in terms of sex (P=.45).

Outcome Variables and Predictors

Cognitive impairment was defined by the Chinese version of the Mini-Mental State Examination (CMMSE) [24], which was culturally translated from the international standard of the MMSE questionnaire. The CMMSE contains 24 items within six dimensions (five items for orientation, three for registration, one for naming, five for attention and calculation, three for recall, and seven for language). The score of the Chinese MMSE ranges from 0 to 30 points, with higher scores indicating better cognition. The CMMSE has been validated among the Chinese elderly population, and a score below 18 points has been defined as cognitive impairment [24].

Predictors related to cognitive impairment were assessed a priori based on clinical importance, scientific knowledge, and predictors identified in previously published studies [25]. We therefore selected 45 factors related to demographic characteristics, which included lifestyle, mental health, leisure activities, sleep, chronic diseases, physical function, anthropometric index, and baseline cognitive function (Multimedia Appendix 2).

Statistical Analysis

Categorical variables have been reported as numbers and proportions, and compared using a chi-square test or Fisher exact test. Continuous variables have been expressed as medians with IQRs and compared using the Wilcoxon test when data were not normally distributed. Detailed information is presented in Multimedia Appendix 2. Some covariates contained missing values. The proportion of missing values was less than 5% for all variables. Thus, we performed imputations, using multivariable regression methods via the R package mice. Feature selection was performed using recursive feature elimination (RFE) combined with RF. During the process of elimination, a 10-fold cross-validation was implemented to optimize the variable selection. In addition, the RFE method with the NB method was used to extract variables, and the result was compared with RFE combined with RF. According to the results of RF and NB, the final feature selection was based on the number of features included and accuracy.

We divided the original data into a 2/3 training set, 1/6 validation set, and 1/6 test set [26]. The training set was used for model development. The validation set was used to adjust parameters of the model and explore optimal cutoffs after training was finished. The test set was used to estimate the generalization of the model. Regarding the algorithm used in the development of prediction models, we chose four machine learning algorithms, including RF, XGBoost, NB, and logistic regression, to construct models based on the results of the feature selection. We chose these four learning algorithms because they are recommended by "Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View" [27]. The performances of the four prediction models were compared with each other using areas under the curve, specificity, sensitivity, accuracy, and specificity/sensitivity.

In addition, if logistic regression performs well compared with the other three methods, we will formulate a nomogram based on the result of logistic regression for practical use. The nomogram works by proportionally converting each regression coefficient into a 0 to 100-point scale, with 100 points being the highest β coefficient. The points across each independent variable are added to derive total points, which are translated to predicted probabilities [28]. All analyses were conducted using R, version 3.6.0 [29]. A P value <0.05 was considered to indicate statistical significance.

Results

Population Demographics

A total of 6718 participants were involved. Forty-five explanatory variables were selected, and these variables included nine items of demographic characteristics, five items of lifestyle, 10 items of mental health, five items of leisure activities, two items of sleep, seven items of chronic diseases, two items of physical function, four items of the anthropometric index, and one item of baseline cognitive function. Six variables of demographic characteristics (age, sex, ethnicity, years of education, occupation, and marital status), four variables of lifestyle (fruit, smoking, drinking, and exercise), eight variables of mental health (self-reported quality of life, being positive, hygiene, anxiety, loneliness, decision making, feeling useless, and feeling happy), five variables of leisure activity (garden work, reading, raising pets, playing cards or mah-jongg, and social activities), two variables of sleep (sleep quality and sleep duration), five variables of chronic diseases (hypertension, diabetes, heart disease, cataract, and arthritis), two variables of physical function (activities of daily living and instrumental activities of daily living), one variable of anthropometric measurement (BMI), and baseline MMSE were significantly associated with a 3-year risk of cognitive impairment (P<.001). Detailed information is presented in Multimedia Appendix 2.

Feature Selection

NB combined with RFE showed that accuracy (0.8342) was the highest with four features included in the model (age, instrumental activities of daily living, baseline MMSE, and marital status). RF combined with RFE showed that the model involving 45 variables had the highest accuracy (0.8502), while the model including four variables had an accuracy of 0.8304 (Table 1). Considering the simplicity and accuracy of the prediction model, we finally chose the following four features to develop the model: age, instrumental activities of daily living, baseline MMSE, and marital status.

Table 1. Feature selection using naïve Bayes combined with recursive feature elimination and random forest combined with recursive feature elimination.

Method	Number of features	Accuracy	Kappa	Accuracy SD	Kappa SD
NB[a] combined with RFE[b]	4	0.8342	0.3258	0.007801	0.02452
NB combined with RFE	8	0.8229	0.3543	0.007340	0.02408
NB combined with RFE	16	0.8136	0.3421	0.012724	0.02540
NB combined with RFE	45	0.8315	0.3220	0.007567	0.02639
RF[c] combined with RFE	4	0.8304	0.1545	0.008356	0.05486
RF combined with RFE	8	0.8475	0.1594	0.008545	0.05569
RF combined with RFE	16	0.8471	0.1789	0.007815	0.03612
RF combined with RFE	45	0.8502	0.1214	0.005572	0.04800

[a]NB: naïve Bayes.

[b]RFE: recursive feature elimination.

[c]RF: random forest.

Model Evaluation and Comparison

The training, validation, and test sets involved 4514, 1100, and 1104 points of data, respectively. We tried to use several widely applied machine learning algorithms (RF, NB, XGBoost, and logistic regression) for the construction of the prediction models in the training set.

We used a receiver operating characteristic (ROC) curve, specificity, sensitivity, accuracy, and specificity/sensitivity to evaluate the prediction model in both validation and test data. Before the evaluation, optimal cutoffs were determined by maximizing the Youden index (ie, sensitivity + specificity − 1) by the ROC curve in the validation set. In the test set, ROC curves revealed that logistic regression and NB had better predictive performances, with an area under the curve of 0.814. The area under the curve of XGBoost (0.811) was less than that of logistic regression and NB. RF underperformed, with an area under the curve of 0.780 (Figure 1).

The model of NB performed well in terms of specificity, with a value of 0.776. The specificities of the models of logistic regression (0.770), RF (0.645), and XGBoost (0.738) were lower than that of NB. The model of RF performed well in terms of sensitivity, with a value of 0.793. The sensitivities of the models of logistic regression (0.701), NB (0.672), and XGBoost (0.724) were lower than that of RF. The accuracy of the NB model (0.760) was higher than the accuracies of the other three models. All details about the parameters of the models developed with different algorithms are shown in Table 2.

Figure 1. Receiver operating characteristic curve performance of four models on the test set. AUC: area under the curve.

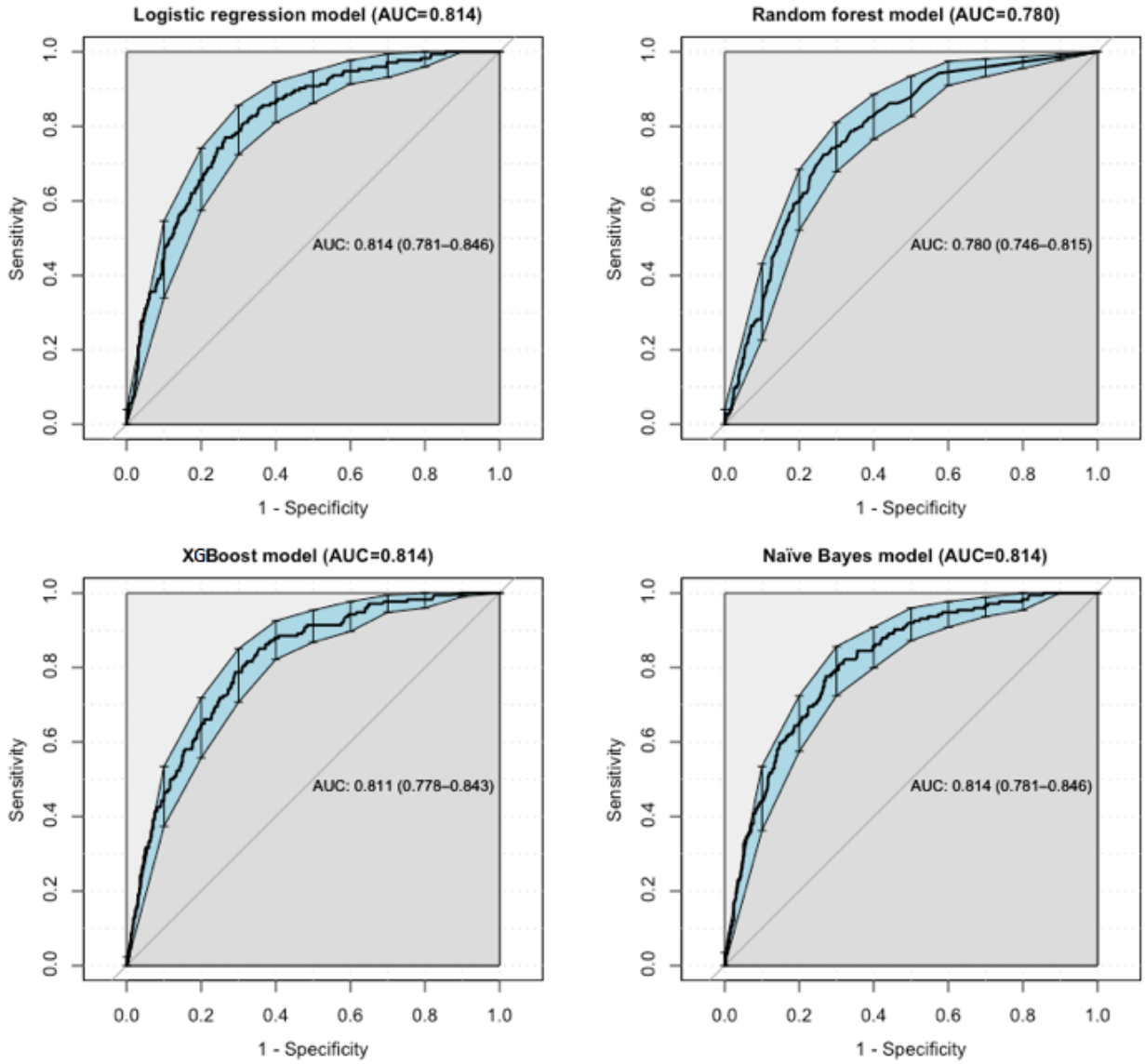


Table 2. Evaluation of the performance of the four algorithms.

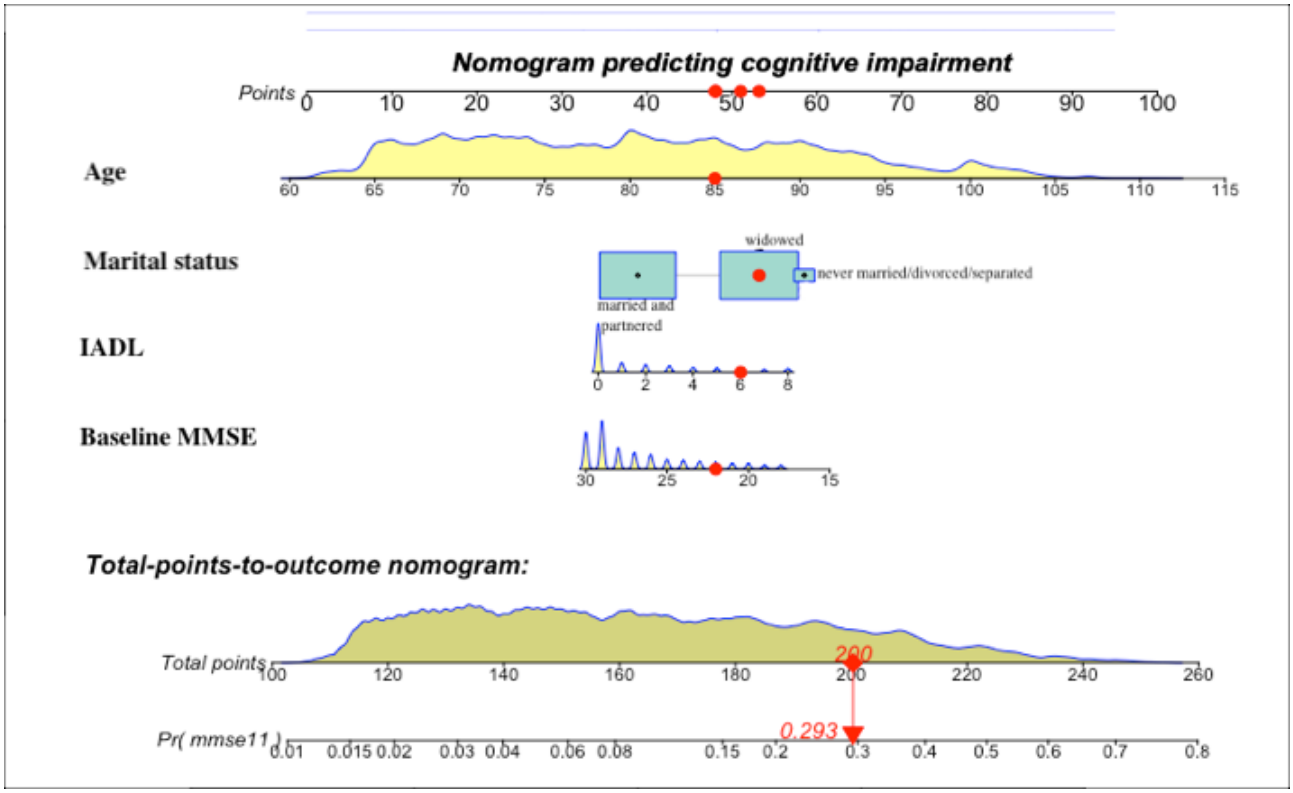
Algorithm	Data set	Area under the curve	Optimal cut-off	Specificity	Sensitivity	Accuracy	Specificity/sensitivity
Logistic regression	Validation	0.812	0.116	0.785	0.682	0.768	1.151
Logistic regression	Test	0.814	0.116	0.770	0.701	0.759	1.098
Random forest	Validation	0.773	0.040	0.654	0.784	0.675	0.834
Random forest	Test	0.780	0.040	0.645	0.793	0.669	0.813
Naïve Bayes	Validation	0.804	0.214	0.796	0.688	0.778	1.157
Naïve Bayes	Test	0.814	0.214	0.776	0.672	0.760	1.155
XGBoost	Validation	0.815	0.302	0.753	0.744	0.752	1.012
XGBoost	Test	0.814	0.302	0.738	0.724	0.736	1.019

Development of the Nomogram

As the prediction developed by logistic regression performed well, a nomogram (Figure 2) was used to present the data and predicted probabilities vividly. The optimal cutoff value of the total nomogram scores was determined to be 170 in the test set. In Figure 2, a box has been used to represent the proportion of the sample in category variables. For instance, for marital status, most of the participants were grouped into “married or partnered” and “widowed.” Density exhibited the distribution

of the sample in continuous variables. For example, for baseline MMSE, most individuals scored over 28. The total points corresponded to the predicted probability. For example, an individual score of 200 points in total, with an age score of 85, a marital status of widowed, an instrumental activities of daily living score of 6, and a baseline MMSE score of 22 corresponded to a predicted probability of cognitive impairment of 29.3%.

Figure 2. Developed nomogram with age, marital status, instrumental activities of daily living (IADL), and baseline Mini-Mental State Examination (MMSE) incorporated.



Discussion

Main Findings

In this study, we developed a prediction model for forecasting the 3-year risk of cognitive impairment among 6718 community-dwelling elderly individuals aged 60 years or older with normal cognitive function via machine learning algorithms. The model performed comparably to the best available biomarkers, such as apolipoprotein genotype [13,17], but is less expensive, easier to implement, and validated internally with reasonable results.

Feature or variable selection is central in the development of a prediction model [30]. Out of the original 45 variables, the following four variables used showed the highest accuracy of the model: age, marital status, instrumental activities of daily living, and baseline MMSE. Our findings support the previous literature regarding age [5,31], marital status [32,33], activities of daily living [34,35], and cognitive status [8,36], indicating that the predictors we selected were usable and reliable.

In this study, age was one of the predictors for the risk of cognitive impairment. It has already been estimated that, in 2020, 83% of people aged 75 years or above live with dementia in the US [37]. Those aged 65 to 74 years account for 17% of cases, while those younger than 65 years could develop dementia, but it is much less common and the prevalence is uncertain [37]. It is indicated that people aged 75 years or above should be screened for the risk of cognitive impairment. However, screening all older people for dementia is not recommended, as the benefits for that are still unclear [5]. Using large existing registers and databases can offer opportunities to explore existing information to predict the health status [38].

We found that marital status was a strong predictor for the risk of cognitive impairment. A recent meta-analysis involving 812,047 participants evaluated the association between marriage and dementia. The results showed that life-long single and widowed older people were respectively 1.42 and 1.20 times more likely to be diagnosed with dementia compared with married older people [33]. Another 10-year longitudinal population-based study including 2,288,489 individuals explored the influence of marital status and concluded that the risk of dementia in nonmarried individuals was around 1.7 times greater than that in married individuals [32]. The results might be explained by the fact that married individuals are more likely to have healthy lifestyles and participate in social activities, which might be conducive to cognitive reserve and reduced dementia risk over a lifespan [39]. On the contrary, those in widowhood might be more likely to experience a higher risk of cognitive impairment than divorced people because of the detrimental effect of stress from bereavement on hippocampal neurons or cognition. Further, as marriage has been considered a social norm, people with difficulties in communication and smaller cognitive reserves across life may be less likely to marry [33]. In today's society, staying unmarried has become more

common, and this phenomenon deserves more attention. Social factors like marital status should be taken seriously as risk determinants for cognitive impairment.

An association between physical function and cognitive capacity among older people has been found in previous studies [34,40]. Our study only included instrumental activities of daily living as a predictor for constructing the model, and it did not consider activities of daily living that represent functional ability. The reason might be that there is a natural hierarchy of functional loss associated with cognitive decline among older people [34]. Older persons with progressive cognitive decline lose the ability to perform tasks, often in the order of bathing, dressing, toileting, transferring, continence, and feeding. Therefore, older people who are not able to feed themselves might not be able to perform other tasks independently [35]. Similarly, a study by Njegovan et al found that among the 14 items of activities of daily living and instrumental activities of daily living, a hierarchy of functional items existed, with instrumental activities of daily living (such as shopping, banking, etc) being lost at higher cognitive scores than basic activities of daily living (such as eating, dressing, etc), which were lost later [34]. Our results confirmed that there was a tendency for instrumental activities of daily living to be a stronger predictor compared with activities of daily living. However, since there was overlap, subdomains of these two tools might be more meaningful for developing a prediction model. For nurses and caregivers, this information can help anticipate the need for intervention in people with declining cognition showing subtle declines in instrumental activities of daily living, which could improve the quality of life of these people and their caregivers and play an important part in health care planning [34].

Baseline cognitive function affected the degree to which cognitive scores changed over time and had a profound effect on further cognitive impairment. In one study using UK biobank data to assess the effect of baseline cognitive performance on a prediction model for 3 to 8-year risk of dementia, the results showed that cognitive performance added up to 5% (from 0.78 to 0.83) to the discriminative accuracy of the ROC model developed with the variables of age, sex, education, family history, and depression [36]. The MMSE has been the most common cognitive variable for developing a dementia prediction model [8]. However, variables of specific cognitive domains, such as memory and executive function, might be more feasible and useful predictors in constructing cognitive impairment prediction models. The total MMSE score was associated less strongly with dementia and Alzheimer disease than the episodic memory subset [8]. Therefore, future studies could consider more specific cognitive domain variables.

Limitations

Our study has limitations, which should be considered in the interpretation of the study results. First, retained cohort members were younger and had on average better cognitive and physical functioning than those who dropped out, which can lead to studies being severely underpowered and biased toward the healthier part of the aging population. As we used a nationally representative database, the ascertainment bias could, to an extent, be limited. Second, we utilized a cross-validation approach to model development and assessment. The results still need to be validated in an independent cohort. Third, we used cohort studies with insufficient details on the duration of marriage, widowhood, and divorce to allow the exploration of a dose-response effect. A future study could take the dose-response effect of marital status on cognitive function into consideration. Fourth, the baseline MMSE was used as one of the predictors. However, other specific cognitive domains, such as memory and executive function, could also be used as features, and they might perform better than whole cognitive function. Fifth, our models were based on a prospective cohort that may have some level of bias. A prospective external validation cohort is needed for further confirmation in future research. Lastly, some predictors used in our study were measured by self-reporting, resulting in information bias. Nevertheless, self-reported data are more feasibly collected in primary health care settings, and the results can be generalized to wider communities. Despite these limitations, we believe that the results are usable in terms of cognitive impairment prevention and further intervention globally.

Implications

This is one of the first studies where a machine learning approach has been used in a nursing context. The study showed that machine learning can be used more widely in nursing science in different contexts and various functions. The prediction models exert implications in the three-grade prevention system of diseases [38]. In the primary prevention of diseases, a cognitive impairment prediction model could provide quantitative risk value (probability) of cognitive deficiency in the next 3 years, based on the current health status, offering a more intuitive and powerful scientific tool for health education and behavioral intervention. In the secondary prevention of diseases, using noninvasive, low-cost, and easy-to-acquire variables to develop a prediction model is more practicable for staff, particularly general practitioners in community health, to bring about “early detection, early diagnosis, and early treatment,” which have large influences on medical costs for dementia. In the tertiary prevention of diseases, the prediction model could be used to predict recurrence, reducing mortality and disability [38]. A simple and feasible prediction model would also help nurses to be aware of the progression of diseases over time. Therefore, nurses could be better aware of triggers that might alarm them about any hidden problems. In addition, a precise prediction model with predictors that are more available in clinical environments could help clinical nurses understand the prognostic factors of diseases. Based on this information, nurses could offer tailored preventive interventions to patients before any signs of cognition deficits occur.

This study provides guidance for future research as well. First, the use of several algorithms to construct prediction models in specific diseases offers more opportunity to find a more suitable model with a high area under the curve and accuracy. Second, selecting the most suitable predictors is important for developing a prediction model to use in clinical practice. Easy-to-acquire, noninvasive, and low-cost variables are welcome in clinical nursing, and invasive biomarkers could improve the prediction. The former is more suitable for community health care and any

clinical environment because of large populations and insufficient staff and funds, while the latter is more applicable in more specific clinical environments for people with high risk of diseases. Lastly, we included Chinese elderly people aged 60 years or above and developed a cognitive impairment prediction model. Further studies could develop cognitive impairment prediction models for middle-aged people as the World Health Organization has suggested to increase the cognitive reserve in mid-life and early aging (45-70 years) [41].

In the future, the results of this study could be used in countries and areas with less human resources, such as low- and middle-income countries, to identify elderly people with a high risk of developing cognitive deficiency in the next 3 years (ie, age, marital status, physical function, and cognitive function). Simple, relevant, and easy-to-detect risk factors would save time and resources in health care and would especially help nursing staff identify those people who are at high risk of developing cognitive impairment. As family members living with elderly people do not always recognize the early signs of dementia [42], the knowledge obtained from this study could be used to educate family members as well.

Acknowledgments

We appreciate the support from the Special Funding for the Construction of Innovative Provinces in Hunan (grant number: 2019SK2141), China Scholarship Council (grant number: Liumeijin 2020-1509), and Central South University-Flinders University Seed Project (grant number: 2020xyhlzzjj002).

Authors' Contributions

MYH and XYW conceived of the presented idea. MYH and XHS performed the analytical calculations and took the lead in writing the manuscript. GY verified the analytical methods. VM investigated and supervised the findings of this work and helped in the language edit. HF supervised the whole project. In addition, VM and HF contributed equally to this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flow chart of participant selection.
[DOCX File , 25 KB - jmir_v23i2e20298_app1.docx]

Multimedia Appendix 2
Detailed information about 45 features among the participants (grouped by cognitive impairment).
[DOCX File , 35 KB - jmir_v23i2e20298_app2.docx]

References

1. World Alzheimer Report 2015: The global impact of dementia: an analysis of prevalence, incidence, cost and trends. Alzheimer's Disease International. 2015. URL: https://www.alzint.org/resource/world-alzheimer-report-2015/ [accessed 2021-02-12]
2. World Alzheimer Report 2019: Attitudes to dementia. Alzheimer's Disease International. 2019. URL: https://www.alzint.org/resource/world-alzheimer-report-2019/ [accessed 2021-01-12]
3. Frankish H, Horton R. Prevention and management of dementia: a priority for public health. The Lancet 2017 Dec;390(10113):2614-2615. [doi: 10.1016/s0140-6736(17)31756-7]
4. World Alzheimer Report 2018-The state of the art of dementia research: New frontiers. Alzheimer's Disease International. 2018. URL: https://www.alzint.org/u/WorldAlzheimerReport2018.pdf [accessed 2021-01-12]
5. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. The Lancet 2017 Dec;390(10113):2673-2734. [doi: 10.1016/s0140-6736(17)31363-6]
6. Chen X, Zhou Y, Wang R, Cao H, Reid S, Gao R, Alzheimer's Disease Neuroimaging Initiative. Potential Clinical Value of Multiparametric PET in the Prediction of Alzheimer's Disease Progression. PLoS One 2016 May 16;11(5):e0154406 [FREE Full text] [doi: 10.1371/journal.pone.0154406] [Medline: 27183116]
7. Kivipelto M, Ngandu T, Laatikainen T, Winblad B, Soininen H, Tuomilehto J. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. The Lancet Neurology 2006 Sep;5(9):735-741. [doi: 10.1016/s1474-4422(06)70537-3]
8. Hou X, Feng L, Zhang C, Cao X, Tan L, Yu J. Models for predicting risk of dementia: a systematic review. J Neurol Neurosurg Psychiatry 2019 Apr 28;90(4):373-379. [doi: 10.1136/jnnp-2018-318212] [Medline: 29954871]
9. Pase MP, Beiser AS, Himali JJ, Satizabal CL, Aparicio HJ, DeCarli C, et al. Assessment of Plasma Total Tau Level as a Predictive Biomarker for Dementia and Related Endophenotypes. JAMA Neurol 2019 May 01;76(5):598-606 [FREE Full text] [doi: 10.1001/jamaneurol.2018.4666] [Medline: 30830207]

10. Ding D, Zhao Q, Guo Q, Liang X, Luo J, Yu L, Shanghai Aging Study (SAS). Progression and predictors of mild cognitive impairment in Chinese elderly: A prospective follow-up in the Shanghai Aging Study. Alzheimers Dement (Amst) 2016;4:28-36 [FREE Full text] [doi: 10.1016/j.dadm.2016.03.004] [Medline: 27489876]
11. Bratić B, Kurbalija V, Ivanović M, Oder I, Bosnić Z. Machine Learning for Predicting Cognitive Diseases: Methods, Data Sources and Risk Factors. J Med Syst 2018 Oct 27;42(12):243. [doi: 10.1007/s10916-018-1071-x] [Medline: 30368611]
12. Joshi S, Deepa Shenoy P, Simha GG V, Venugopal KR, Patnaik L. Classification of Neurodegenerative Disorders Based on Major Risk Factors Employing Machine Learning Techniques. IJET 2010;2(4):350-355. [doi: 10.7763/ijet.2010.v2.146]
13. Ford E, Rooney P, Oliver S, Hoile R, Hurley P, Banerjee S, et al. Identifying undetected dementia in UK primary care patients: a retrospective case-control study comparing machine-learning and standard epidemiological approaches. BMC Med Inform Decis Mak 2019 Dec 02;19(1):248 [FREE Full text] [doi: 10.1186/s12911-019-0991-9] [Medline: 31791325]
14. Li R. Data Mining and Machine Learning Methods for Dementia Research. In: Perneczky R, editor. Biomarkers for Alzheimer's Disease Drug Development. Methods in Molecular Biology, vol 1750. New York, NY: Humana Press; 2018:363-370.
15. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. The Lancet Psychiatry 2016 Mar;3(3):243-250. [doi: 10.1016/s2215-0366(15)00471-x]
16. Cleret de Langavant L, Bayen E, Yaffe K. Unsupervised Machine Learning to Identify High Likelihood of Dementia in Population-Based Surveys: Development and Validation Study. J Med Internet Res 2018 Jul 09;20(7):e10493 [FREE Full text] [doi: 10.2196/10493] [Medline: 29986849]
17. Hall A, Pekkala T, Polvikoski T, van Gils M, Kivipelto M, Lötjönen J, et al. Prediction models for dementia and neuropathology in the oldest old: the Vantaa 85+ cohort study. Alzheimers Res Ther 2019 Jan 22;11(1):11 [FREE Full text] [doi: 10.1186/s13195-018-0450-3] [Medline: 30670070]
18. Pekkala T, Hall A, Lötjönen J, Mattila J, Soininen H, Ngandu T, et al. Development of a Late-Life Dementia Prediction Index with Supervised Machine Learning in the Population-Based CAIDE Study. JAD 2016 Dec 06;55(3):1055-1067. [doi: 10.3233/jad-160560]
19. Xie Y, Yu Z. Models and prediction, how and what? Ann Transl Med 2020 Feb;8(4):75-75 [FREE Full text] [doi: 10.21037/atm.2019.12.133] [Medline: 32175368]
20. Obermeyer Z, Emanuel EJ. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016 Sep 29;375(13):1216-1219. [doi: 10.1056/nejmp1606181]
21. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD). Circulation 2015 Jan 13;131(2):211-219. [doi: 10.1161/circulationaha.114.014508]
22. Zeng Y. Chinese Longitudinal Healthy Longevity Survey and some research findings. Geriatr Gerontol Int 2004 Sep;4(s1):S49-S52. [doi: 10.1111/j.1447-0594.2004.00147.x]
23. Center for Healthy Aging and Development Studies. Peking University Open Research Data Platform. URL: https://opendata.pku.edu.cn/dataverse/CHADS [accessed 2021-02-11]
24. Zhang Z, Zahner GEP, Román GC, Liu X, Wu C, Hong Z, et al. Socio-demographic variation of dementia subtypes in china: Methodology and results of a prevalence study in Beijing, Chengdu, Shanghai, and Xian. Neuroepidemiology 2006;27(4):177-187 [FREE Full text] [doi: 10.1159/000096131] [Medline: 17035714]
25. Kim Y, Margonis GA, Prescott JD, Tran TB, Postlewait LM, Maithel SK, et al. Nomograms to Predict Recurrence-Free and Overall Survival After Curative Resection of Adrenocortical Carcinoma. JAMA Surg 2016 Apr 01;151(4):365-373 [FREE Full text] [doi: 10.1001/jamasurg.2015.4516] [Medline: 26676603]
26. Lee S, Gui X, Manquen M, Hamilton E. Use of Training, Validation, and Test Sets for Developing Automated Classifiers in Quantitative Ethnography. In: Eagan B, Misfeldt M, Siebert-Evenstone A, editors. Advances in Quantitative Ethnography. ICQE 2019. Communications in Computer and Information Science, vol 1112. Cham: Springer; 2019:117-127.
27. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. J Med Internet Res 2016 Dec 16;18(12):e323 [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]
28. Lei Z, Li J, Wu D, Xia Y, Wang Q, Si A, et al. Nomogram for Preoperative Estimation of Microvascular Invasion Risk in Hepatitis B Virus-Related Hepatocellular Carcinoma Within the Milan Criteria. JAMA Surg 2016 Apr 01;151(4):356-363. [doi: 10.1001/jamasurg.2015.4257] [Medline: 26579636]
29. The R Project for Statistical Computing. URL: http://www.r-project.org/ [accessed 2021-02-11]
30. Deo RC. Machine Learning in Medicine. Circulation 2015 Nov 17;132(20):1920-1930. [doi: 10.1161/circulationaha.115.001593]
31. Chinese Dementia Cognitive Impairment Diagnosis Treatment Guidelines Writing Group. Diagnosis and treatment of dementia (2018): risk factors and interventions of Alzhemer's disease. Natl Med J China 2018;98(19):1461-1466.
32. Sundström A, Westerlund O, Kotyrlo E. Marital status and risk of dementia: a nationwide population-based prospective study from Sweden. BMJ Open 2016 Jan 04;6(1):e008565 [FREE Full text] [doi: 10.1136/bmjopen-2015-008565] [Medline: 26729377]

33. Sommerlad A, Ruegger J, Singh-Manoux A, Lewis G, Livingston G. Marriage and risk of dementia: systematic review and meta-analysis of observational studies. J Neurol Neurosurg Psychiatry 2018 Mar 28;89(3):231-238 [FREE Full text] [doi: 10.1136/jnnp-2017-316274] [Medline: 29183957]
34. Njegovan V, Hing MM, Mitchell SL, Molnar FJ. The hierarchy of functional loss associated with cognitive decline in older persons. J Gerontol A Biol Sci Med Sci 2001 Oct 01;56(10):M638-M643. [doi: 10.1093/gerona/56.10.m638] [Medline: 11584037]
35. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. JAMA 1963 Sep 21;185:914-919. [doi: 10.1001/jama.1963.03060120024016] [Medline: 14044222]
36. Calvin CM, Wilkinson T, Starr JM, Sudlow C, Hagenaars SP, Harris SE, et al. Predicting incident dementia 3-8 years after brief cognitive tests in the UK Biobank prospective study of 500,000 people. Alzheimers Dement 2019 Dec 13;15(12):1546-1557. [doi: 10.1016/j.jalz.2019.07.014] [Medline: 31619348]
37. Hebert LE, Weuve J, Scherr PA, Evans DA. Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. Neurology 2013 Feb 06;80(19):1778-1783. [doi: 10.1212/wnl.0b013e31828726f5]
38. Zhou Z, Wang W, Li Y, Jin K, Wang X, Wang Z, et al. In-depth mining of clinical data: the construction of clinical prediction model with R. Ann Transl Med 2019 Dec;7(23):796-796 [FREE Full text] [doi: 10.21037/atm.2019.08.63] [Medline: 32042812]
39. Kuiper JS, Zuidersma M, Oude Voshaar RC, Zuidema SU, van den Heuvel ER, Stolk RP, et al. Social relationships and risk of dementia: A systematic review and meta-analysis of longitudinal cohort studies. Ageing Res Rev 2015 Jul;22:39-57. [doi: 10.1016/j.arr.2015.04.006] [Medline: 25956016]
40. Barberger-Gateau P, Commenges D, Gagnon M, Letenneur L, Sauvel C, Dartigues J. Instrumental activities of daily living as a screening tool for cognitive impairment and dementia in elderly community dwellers. J Am Geriatr Soc 1992 Nov 27;40(11):1129-1134. [doi: 10.1111/j.1532-5415.1992.tb01802.x] [Medline: 1401698]
41. Vellas B, Scrase D, Rosenberg G, Andrieu S, Araujo de Carvalho I, Middleton L. Editorial: WHO Guidelines on Community-Level Interventions to Manage Declines in Intrinsic Capacity: The Road for Preventing Cognitive Declines in Older Age? J Prev Alzheimers Dis 2018;5(3):165-167. [doi: 10.14283/jpad.2018.26] [Medline: 29972207]
42. World Alzheimer Report 2016-Improving healthcare for people living with dementia: Coverage, quality and costs now and in the future. Alzheimer's Disease International. 2016. URL: https://www.alzint.org/u/WorldAlzheimerReport2016.pdf [accessed 2021-01-12]

Abbreviations

CLHLS: Chinese Longitudinal Healthy Longevity Survey
CMMSE: Chinese version of the Mini-Mental State Examination
MMSE: Mini-Mental State Examination
NB: naïve Bayes
RF: random forest
RFE: recursive feature elimination
ROC: receiver operating characteristic
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

Edited by G Eysenbach, Q Zeng; submitted 15.05.20; peer-reviewed by X Ling, C Li; comments to author 28.08.20; revised version received 14.01.21; accepted 18.01.21; published 24.02.21.

Please cite as:
Hu M, Shu X, Yu G, Wu X, Välimäki M, Feng H
A Risk Prediction Model Based on Machine Learning for Cognitive Impairment Among Chinese Community-Dwelling Elderly People With Normal Cognition: Development and Validation Study
J Med Internet Res 2021;23(2):e20298
URL: https://www.jmir.org/2021/2/e20298
doi:10.2196/20298
PMID:33625369

Original Paper

Conversational Agents as Mediating Social Actors in Chronic Disease Management Involving Health Care Professionals, Patients, and Family Members: Multisite Single-Arm Feasibility Study

Tobias Kowatsch[1,2,3], PhD; Theresa Schachner[1], BA, BSc, MSc; Samira Harperink[1], BA; Filipe Barata[1], MSc, PhD; Ullrich Dittler[4], Prof Dr; Grace Xiao[5], BA; Catherine Stanger[6], PhD; Florian v Wangenheim[1,2], Prof Dr; Elgar Fleisch[1,2,3], PhD; Helmut Oswald[7], MD; Alexander Möller[8], MD, Prof Dr

[1]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[2]Future Health Technologies Programme, Campus for Research Excellence and Technological Enterprise, Singapore-ETH Centre, Singapore, Singapore

[3]Centre for Digital Health Interventions, Institute of Technology Management, University of St Gallen, St Gallen, Switzerland

[4]Fakultät Digitale Medien, Campus Furtwangen, Hochschule Furtwangen University, Furtwangen, Germany

[5]Johns Hopkins University School of Medicine, Johns Hopkins University, Baltimore, MD, United States

[6]Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States

[7]Department of Child and Adolescent Health, Cantonal Hospital Winterthur, Winterthur, Switzerland

[8]Division of Respiratory Medicine and Childhood Research Center, University Children's Hospital Zurich, Zurich, Switzerland

Corresponding Author:
Tobias Kowatsch, PhD
Centre for Digital Health Interventions
Department of Management, Technology, and Economics
ETH Zurich
Weinbergstrasse 56/58
Zurich, 8092
Switzerland
Phone: 41 71 224 72 44
Email: tkowatsch@ethz.ch

Abstract

Background: Successful management of chronic diseases requires a trustful collaboration between health care professionals, patients, and family members. Scalable conversational agents, designed to assist health care professionals, may play a significant role in supporting this collaboration in a scalable way by reaching out to the everyday lives of patients and their family members. However, to date, it remains unclear whether conversational agents, in such a role, would be accepted and whether they can support this multistakeholder collaboration.

Objective: With asthma in children representing a relevant target of chronic disease management, this study had the following objectives: (1) to describe the design of MAX, a conversational agent–delivered asthma intervention that supports health care professionals targeting child-parent teams in their everyday lives; and (2) to assess the (a) reach of MAX, (b) conversational agent–patient working alliance, (c) acceptance of MAX, (d) intervention completion rate, (e) cognitive and behavioral outcomes, and (f) human effort and responsiveness of health care professionals in primary and secondary care settings.

Methods: MAX was designed to increase cognitive skills (ie, knowledge about asthma) and behavioral skills (ie, inhalation technique) in 10-15-year-olds with asthma, and enables support by a health professional and a family member. To this end, three design goals guided the development: (1) to build a conversational agent–patient working alliance; (2) to offer hybrid (human- and conversational agent–supported) ubiquitous coaching; and (3) to provide an intervention with high experiential value. An interdisciplinary team of computer scientists, asthma experts, and young patients with their parents developed the intervention collaboratively. The conversational agent communicates with health care professionals via email, with patients via a mobile chat app, and with a family member via SMS text messaging. A single-arm feasibility study in primary and secondary care settings was performed to assess MAX.

Results: Results indicated an overall positive evaluation of MAX with respect to its reach (49.5%, 49/99 of recruited and eligible patient-family member teams participated), a strong patient-conversational agent working alliance, and high acceptance by all relevant stakeholders. Moreover, MAX led to improved cognitive and behavioral skills and an intervention completion rate of 75.5%. Family members supported the patients in 269 out of 275 (97.8%) coaching sessions. Most of the conversational turns (99.5%) were conducted between patients and the conversational agent as opposed to between patients and health care professionals, thus indicating the scalability of MAX. In addition, it took health care professionals less than 4 minutes to assess the inhalation technique and 3 days to deliver related feedback to the patients. Several suggestions for improvement were made.

Conclusions: This study provides the first evidence that conversational agents, designed as mediating social actors involving health care professionals, patients, and family members, are not only accepted in such a "team player" role but also show potential to improve health-relevant outcomes in chronic disease management.

(J Med Internet Res 2021;23(2):e25060) doi:10.2196/25060

KEYWORDS

digital health intervention; intervention design; mHealth; eHealth; chatbot; conversational agent; chronic diseases; asthma; feasibility study

Introduction

Chronic conditions present a significant risk to the global population, and cause substantial financial and health-related burdens, resulting in low quality of life of those affected [1]. Chronic diseases affected more than half the population of the United States in 2016, representing a leading cause of death, and their prevalence is expected to rise even further [1]. In addition to ongoing treatment and medical oversight, disease management is a key pillar for chronic condition alleviation by aiming to minimize their symptoms, resulting functional impairments, and related exacerbating risks [2].

Successful disease management often requires a trustful collaboration among health care professionals, patients, and their families [3]. In addition, patients require specialized cognitive and behavioral skills to deal with their condition [4,5]. This is especially important for affected children who have to deal with their disease for their upcoming future [3,6].

Digital health interventions are emerging tools for the management of chronic diseases, as they can educate and engage patients through a direct channel that supports communication with physicians and health care professionals [7,8], and enable the scale up of personalized and behavioral interventions at low cost [1,9]. Digital health interventions offer medical care outside the clinical setting to provide ongoing support and communication in everyday monitoring and management [1]. Indeed, several recent studies have provided evidence supporting the patient benefits of such digital interventions, particularly in children and adolescents [10-14]. In addition, conversational agents (ie, computer programs that imitate interaction with human beings) have shown promising results with respect to patient satisfaction [15], therapeutic alliance [16,17], and treatment success [18]. Digital health interventions in the form of mobile apps can be particularly effective for children as they provide an attractive channel for education and management through the possible integration of multimedia content such as audio or video [19]. Conversational agents, as part of such interventions, can act as mediating social actors; that is, they take over not only a significant amount of the intervention delivery in a scalable manner but also coordinate the communication among health care professionals, family members, and patients if required.

This study focused on asthma as a representative chronic condition. Affecting approximately 235 million people, asthma is one of the most common chronic diseases worldwide [20]. Asthma is characterized by reversible airway obstruction [21]. Its symptoms include wheezing, shortness of breath, and coughing [22]. Asthma is associated with high financial and health costs, with total annual asthma costs in the United States estimated at US $56 billion in 2011 [23]. Depending on the country, the mean cost of asthma care per patient per year can range from US $1900 in Europe to US $3100 in the United States [23]. Even though lack of medical treatment leads to significantly reduced quality of life, the management of asthma still presents a daunting challenge because the exact cause is not well known and its appearance varies significantly between individuals [24].

For asthma, specific cognitive skills required for disease management include knowledge about asthma triggers and the importance of medication inhalation adherence, as well as behavioral skills such as the application of correct inhalation techniques. Further, asthma education and health literacy are fundamental to self-management since better understanding of the condition would help patients avoid the negative effects of poor asthma control [25-27]. Low levels of health literacy have been linked with adverse health outcomes, including more frequent hospitalization and longer stays, even after controlling for severity of illness and socioeconomic variables [28,29].

However, young patients still face problems related to both cognitive and behavioral skills that hinder their ability to effectively administer asthma medications [30-35]. For example, knowledge about asthma or important facets of asthma control such as the importance of medication adherence might change over time, making it necessary for patients to continuously update their knowledge base [36-40]. Another common concern is poor technique during medication inhalation, leading to reduced dispersion of the drug in the lungs, and subsequent decreased asthma stability and lowered clinical effectiveness of the delivered drug [41-44].

Numerous mobile apps have been developed for the management of asthma with particular focus on tracking symptoms or medications [45]. Asthma apps targeted at children often include a gamification component to increase engagement, and to familiarize them with aspects of asthma monitoring and management such as medication intake [46,47]. However, and in addition to shortcomings of asthma management related to cognitive and behavioral skills, children often face problems with such technological solutions when they are not integrated into existing health care systems and do not allow for explicit support by health care professionals or family members. Without a dedicated party or mediator, it often becomes a challenge to integrate all of these relevant stakeholders (ie, health care professionals, family members, and the patients themselves) into the disease management process.

Additionally, due to absent or insufficient motivation strategies such as interactivity, proper incentives and rewards [48], and experiential value [49-51], the effects of the previously reported digital interventions in asthma, such as the health condition of the young patients, are prone to be negatively affected by the temporal decline in the patients' engagement and motivation [10,52-54]. A patient's motivation to comply with digital interventions and adhere to therapeutic tasks may be further diminished by various factors such as family routines; child-raising issues; social issues [55]; and trust, communication, and empathy with health care professionals [56]. Moreover, there is evidence that shared decision making and collaboration among patients, parents, and health care professionals are key success factors in guided asthma self-management programs with improved adherence and health outcomes [55].

Against this background, our research questions are (1) whether conversational agents would be generally adopted for developing a trustful collaboration among health care professionals, young patients, and their family; and (2) whether they could have a positive impact on the management of asthma in children. To answer these questions, this study had the following objectives: (1) to describe the design of MAX, a conversational agent–delivered asthma intervention that supports health care professionals targeting children-parent teams in their everyday lives; and (2) to assess the (a) reach of MAX, (b) conversational agent–patient working alliance, (c) acceptance of MAX, (d) intervention completion rate, (e) cognitive and behavioral outcomes, and (f) human effort and responsiveness of health care professionals in both primary and secondary care settings.

Methods

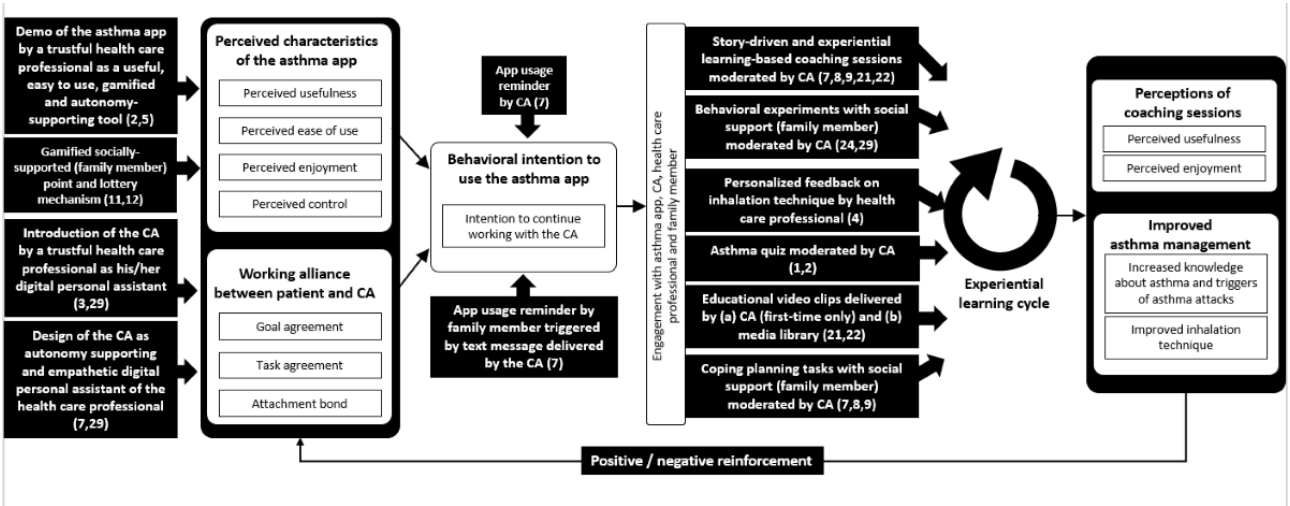
Design

MAX was designed for the German-speaking part of Switzerland, and was evaluated in two home care settings and four secondary care settings at hospitals. The following subsections describe the design and evaluation procedures in detail.

Conceptual Model

Following the preparation phase of the multiphase optimization strategy for behavioral interventions [57], we started with the design of the conceptual model of MAX (see Figure 1). The design of the conceptual model was theoretically informed by related work covering asthma management in children (see Introduction), information systems and technology acceptance research [49-51,58], working alliance [59,60] linked to conversational agents [16,61-64], behavior change techniques (BCTs) [65], and experiential learning theory [66]. Moreover, feedback from four asthma experts of the Swiss Lung Association; two pediatric pneumologists of Swiss children's hospitals; young asthma patients and their parents; and lessons learnt from prior work, in which we developed conversational agents for children with obesity [67,68], were used in the design process. The resulting conceptual model reflects the causal chain triggered by intervention components that target (1) the engagement of young patients with the asthma app, conversational agent, health care professional, and supporting family member (left part of Figure 1), and (2) the outcomes of the intervention (right part of Figure 1).

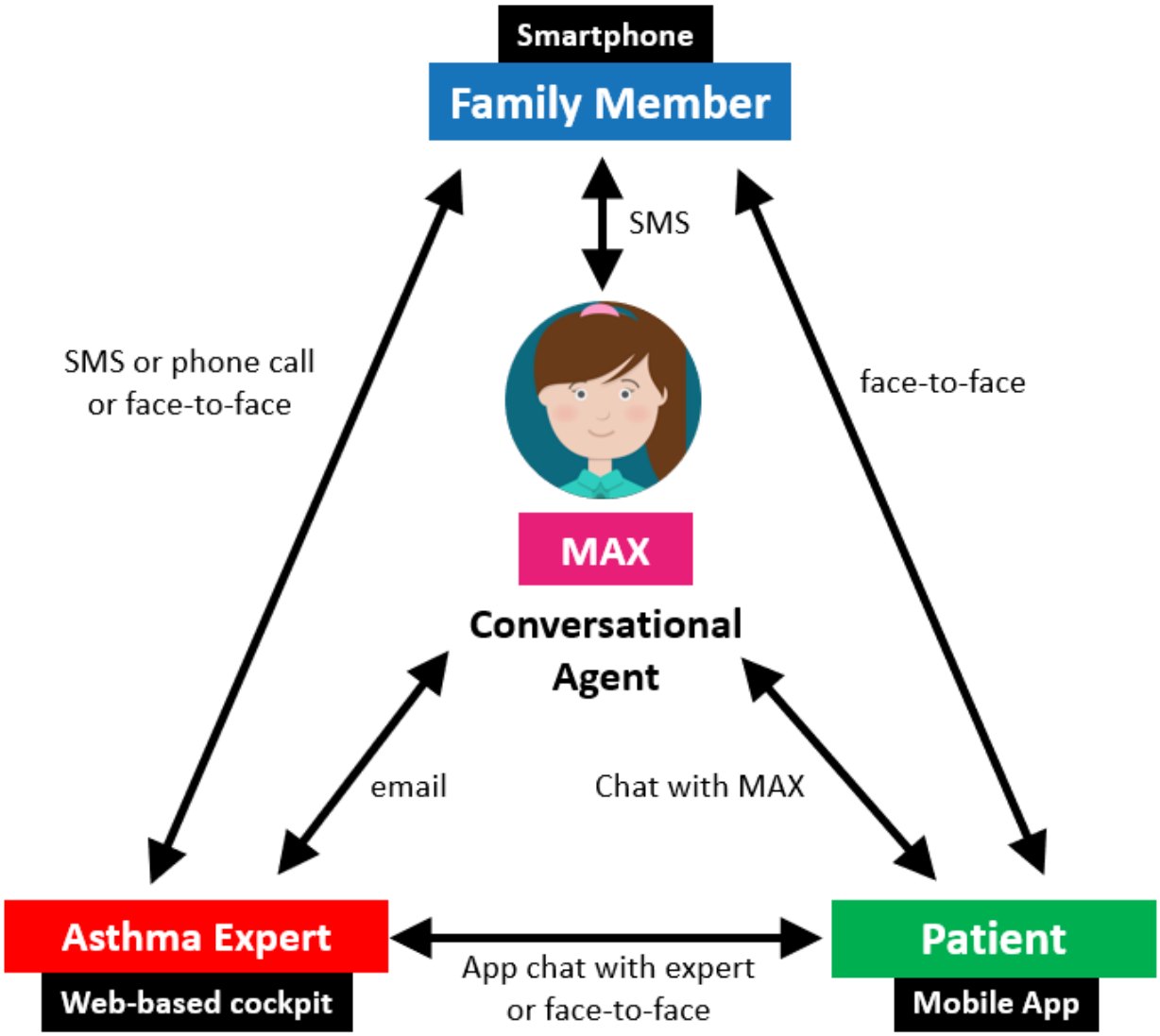
Figure 1. Conceptual model of the intervention. Intervention components are represented by black boxes; behavioral change technique numbers [65] are listed in brackets for each intervention component. CA: conversational agent.



Communication Concept

The communication concept of the intervention allowed patients to engage with the asthma app, the conversational agent MAX, health care professionals, and family members via different communication channels. The communication concept is depicted in Figure 2. In line with self-determination theory [69], which describes autonomy (ie, the need to self-regulate one's experiences and actions as important predictors of engagement [70]), the setup of this communication system allowed patients to independently decide with whom to interact and when, to establish relatedness to all involved stakeholders, and to ultimately increase their competence in the form of improved asthma management.

Figure 2. Communication concept of social actors; MAX combines different communication channels and incorporates family members, patients, and asthma experts into on-site and remote counseling sessions.



The conversational agent itself followed a predefined intervention schedule route (see Multimedia Appendix 1 and 2) to communicate with all participating groups (ie, with health care professionals via email, with patients via a mobile chat app, and with a family member via SMS text messaging; see Figure 2). On top of these channels, there was an on-demand option to communicate via these and the other channels (eg, phone call or face-to-face interaction when required or triggered by parents, the patient, or the health care professional).

Besides the mobile app, the intervention offered a web-based MAX interface for health care professionals (see Multimedia Appendix 3), which was only accessible to the participating health care professionals to interact with their patients when required for a coaching session, monitoring their performance, or accessing their personal information such as patient ID. Patients first accessed the MAX app via a QR code printed on a physical card, which was handed out to them by their health care professional at the beginning of the intervention (see Multimedia Appendix 4). Each participant was linked to a personalized code printed on this card. This connection between the patient and health care professional allowed the conversational agent to recognize the treating health professional so that it can link back to its assigned human as needed.

Intervention Components Triggering Adoption of and Adherence to MAX

To trigger engagement and in line with the theories of planned behavior [71], self-determination theory [69], and technology acceptance research [49,50,58], perceived characteristics of the asthma app (ie, perceived usefulness, ease of use, enjoyment, and control) and working alliance with the conversational agent (ie, goal agreement, task agreement, and attachment bond) were hypothesized to positively influence the behavioral intention to continue working with the conversational agent.

To positively influence the perceived characteristics of the asthma app, health care professionals, who aim to build a trustful relationship with their patients, as this is central in health care situations [72,73], were asked to *demonstrate the app to their patients as a useful, easy to use, gamified, and autonomy-supporting tool* for their asthma management. Toward this end, they provided information on the consequences of behavior to the individual (BCT 2 [65]) and aimed at goal setting (BCT 5 [65]) early on. They introduced patients to the overall communication concept of MAX (see Figure 2), which allowed the integration of all involved stakeholders and the realization of a hybrid ubiquitous coaching approach via on-site and remote counseling sessions. This onboarding activity with health care professionals further aimed at attaining the perception of the conversational agent as a trustworthy social actor that complements the health care professional and family team.

Moreover, a *gamified and socially supported point and lottery mechanism* was implemented as an intervention component to positively influence the patient's perceived enjoyment of the app, which was designed to influence the subsequent behavioral intention of patients to use the asthma app. This intervention component was informed by BCTs 11 and 12 (prompt review of outcome goals and prompt rewards contingent on effort or progress toward behavior, respectively [65]). In detail, patients received 10 points for their first conversational turns with MAX, the participation in a health literacy quiz, and then for each finished coaching session. An additional 10 points were rewarded for each of the seven sessions a family member was involved in and to support the patient. Upon finishing the intervention within 30 days, points accumulated up to that point were doubled. To this end, patients received automatic reminders about how many days they had left for qualifying to double their points during the program. Another 100 points were awarded when the family member completed a final survey at the end of the intervention and handed out the generated code to the patient to withdraw the bonus points. The final points were converted into chances for a lottery. Three winners were drawn from each participating Swiss canton (for more details on the study design see the Study Design section). Each winner received a gift voucher worth US $50 for the Apple App Store, Google Play Store, or a visit to a local movie theatre.

To build up a working alliance, health care professionals were asked to *introduce the conversational agent as their personal digital assistant* (BCTs 3 and 29, provide information about others' approval and plan social support/social change, respectively [65]). In addition, we designed the conversational agent *as an autonomy-supporting* (eg, patients were able to control, and set up a date and time of the digital coaching sessions) *and empathetic digital assistant of the health care professional* (eg, the conversational agent introduced itself as the personal assistant of a health care professional by mentioning his/her name, and, several times during the intervention, the conversational agent asked the patients about their emotional state and provided personalized feedback based on their answers) in accordance with BCTs 7 and 29 (action planning and plan social support/social change, respectively [65]).

Moreover, *app usage reminders* were triggered by the conversational agent as in-app notifications (after 1 hour, 1 day, and 3 days of no interaction) and through a separate communication channel such as via SMS text messaging (after 5 days to the patient and after 7 days to the family member's smartphone) to positively influence the intention of the patient to continue working with the conversational agent. These reminders endorsed action planning (BCT 7 [65]) and further supported the development of relatedness [69] between patients and their parents as important participants of the intervention.

Intervention Components Triggering Experiential Learning and Outcomes

Four distinct intervention components as depicted in Figure 1 enabled an experiential learning cycle [66], and were assumed to influence the outcomes of the intervention, including perception of the coaching sessions (ie, perceived usefulness and perceived enjoyment) and the improvement of individual asthma management (ie, increased knowledge about asthma and triggers of asthma attacks, and improved inhalation technique). Experiential learning describes learning as a process that is continuously grounded in experience and understood as a holistic process that fosters adaptions of the learner to the surrounding reality [66]. The four cyclic steps that describe this process—active experimentation, concrete experience, reflective observation, and abstract conceptualization [66]—are triggered by the intervention components.

The story-driven and experiential learning–based coaching sessions moderated by the conversational agent as the overarching intervention component foster active experimentation [66], concrete experiences [66], and implement several BCTs (7-9, 21, 22; see [65] for a detailed description and Multimedia Appendix 1 and 2 for an overview of the coaching sessions). For patients, coaching sessions were moderated by the conversational agent MAX, which offered a relatively simple chat-based interface with predefined answer options to multiple-choice questions, free-text input (eg, asking for the participant's nickname) or number input fields (eg, asking about the participant's age), and a linguistic style that evokes interpersonal closeness as this is assumed to be positively related to the attachment bond between the patient and conversational agent [59,74]. MAX mimicked the behavior of a real human being chatting by using emojis and some humor to build up a social relationship [75] and working alliance [61] when conversing with patients (see Multimedia Appendix 5 and Multimedia Appendix 6). To address participants' accountability, MAX referred to earlier tasks and activities, and gave positive reinforcement. The conversational agent could also send out personalized messages every other day to initiate

a conversation, in which the dialogue began with a warm greeting, followed by questions about the participants' mood such as "How are you today?"

In total, the intervention consisted of 14 individual coaching sessions, in which the topics were designed to increase cognitive skills (ie, knowledge about asthma) and behavioral skills (ie, inhalation technique). Patients could perform a maximum of one coaching session per day to reduce smartphone addiction [76], where each coaching session was designed to last between 10 to 15 minutes. Several coaching sessions required the aid of the supporting family member, such as to film the patient performing an inhalation (see Multimedia Appendix 7 for a representative video clip). The family member was invited by the MAX conversational agent via a corresponding SMS text message at the time the patient made the appointment for that specific coaching session. A detailed schedule of the intervention with an overview of the coaching sessions is outlined in Multimedia Appendix 1.

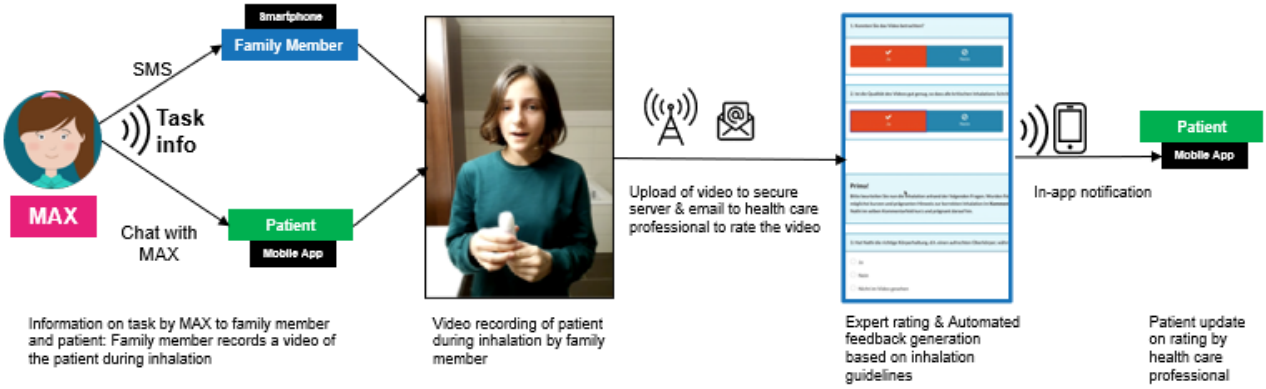
Assuming that the need to self-regulate one's experiences and actions (as an important predictor of engagement [70], as posited by self-determination theory) is also true for digital interventions, the intervention schedule was flexible, which is an innovative approach compared to other interventions [17,68,77-79], and enabled accommodation to the patients' specific needs such as school stress or sickness. Patients could individualize their intervention schedule since they had the option to postpone exercises at their own discretion. This gave patients significant control over the interaction progress and its overall duration. In theory, they could prolong their intervention substantially, but the above-described points reward system incentivized the completion of the program within 30 days by doubling all achieved points when patients complied to this time frame.

The curriculum and storytelling aspects of the intervention were derived from a validated Swiss health literacy comic for children with asthma published by the Swiss Lung Association [80]. Based on this comic, one expert in digital media, didactics, and learning theories wrote a digital health literacy storybook (see Multimedia Appendix 8 and Multimedia Appendix 9), including scripts for 11 health literacy video clips for children with asthma. The storybook was proofread and validated by two asthma experts from the Swiss Lung Association and two pediatric pneumologists. Additionally, established video clips covering correct inhalation techniques for children with asthma were integrated into intervention coaching sessions. These video clips had been produced under the direction of Swiss health care professionals, and are currently used by several Swiss hospitals and patient organizations in their health literacy programs (for links to the video clips see Multimedia Appendix 1 and 2).

Concrete learning experiences [66] were enabled through the intervention component of *behavioral experiments withsocial support (family member) moderated by the conversational agent*. This design allowed patients to relate to the conversational agent and to their social support person. The behavioral experiments addressed asthma management and aimed at improving patients' competence with asthma management [69]. In addition, they enabled environmental restructuring (BCT 24 [65]) and planning of social support/social change (BCT 29 [65]).

During onboarding, health care professionals checked the inclusion criteria with the help of a study recruitment assessment sheet (see Multimedia Appendix 10 and Multimedia Appendix 11) when patients were interested in participating. When patients decided not to participate in the study, health care professionals noted down the corresponding reasons. Further, patients chose their supporting family member and provided their own and their parent's mobile phone numbers to enable communication via the asthma app and mobile phone. Family members provided support to young patients as the intervention component. For example, they were asked to record a short video clip during inhalation as part of a coaching session or filled out a final intervention survey that enabled the young patients to gain more points for the above-described lottery. Figure 3 depicts a representative workflow of the integration of the different stakeholders into the MAX intervention in the course of a behavioral intervention with social support (see Multimedia Appendix 7 for a representative video clip). The family member and the patient are notified over their respective communication channels (ie, SMS and in-app) about an upcoming task. Upon completing the task (in this example, recording the patient during inhalation to evaluate any inhalation mistakes), the conversational agent MAX uploads the video on a secure server and triggers an email notification to the child's health care professional to review the video. The health care professional then assesses the inhalation according to predefined inhalation guidelines (eg, "Did [patient name] exhale enough before inhalation?") with the tags "correct," "not correct," or "not visible in the video." According to these assessments, an automated feedback message is generated, which could be personalized by the health care professional. In a last step, the health care professional sends the personalized feedback message via the web-based MAX interface for health care professionals and the patient receives it as an in-app notification in a separate "health care professional" chat channel. Depending on the severity of the inhalation mistakes, as indicated by the health care professional with an additional yes/no tag, the MAX conversational agent would ask the patient and supporting family member to redo the video recording of the inhalation technique at the beginning of the next coaching session. The goal of this session is to improve the current workflow by extending the reach of health care professionals into the everyday lives of patients in an efficient way without compromising the quality of care and working alliance. That is, parents do not have to arrange a corresponding on-site consultation and travel to the hospital, while at the same time, a standardized assessment of the inhalation video clips as outlined above may even increase the quality of the feedback, and with it the working alliance between the patient and health care provider.

Figure 3. The MAX intervention integrates patients, family members, and health care professionals, and allows a ubiquitous experiential learning experience. An example of session 1 is shown here, in which patients were prompted by the MAX conversational agent to record a video of themselves during inhalation with the help of their support family member, who were additionally informed about the task via SMS text message. Once patients had created and uploaded the video to a secure server, health care professionals received an email to assess the video clip with regard to inhalation mistakes. Patients received their final feedback with comments via in-app notification. See Multimedia Appendix 7 for a representative video clip.



The intervention components *personalized feedback on inhalation technique by the health care professional*, *asthma quiz moderated by the conversational agent*, and *educational video clipsdelivered by (a) the conversational agent (first time only) and (b) media library* allowed for reflective observation [66]. In particular, health care professionals assessed the inhalation technique based on video clips recorded by a patient's family member, provided individual feedback to the patients, and provided normative information about others' behaviors (BCT 4 [65]). This was done via a dedicated chat channel in the web-based MAX interface for health care professionals and the mobile MAX app (see Figure 3), and during on-site visits. This interaction setup extended the dyadic interaction between the patient and conversational agent, resulting in a ubiquitous experiential learning experience besides fostering the relatedness between patients and health care professionals as relevant interaction partners [69].

Health care quizzes were an integral part of the intervention. Patients took a health care quiz at the beginning and end of the intervention as well as short quizzes that were integrated into the conversational turns with the conversational agent MAX. These elements of gamification aimed to increase cognitive skills and provided information on consequences of behavior, both in general as well as to the individual (BCT 1 and 2 [65]). Participants could choose between multiple answers and received feedback depending on the accuracy of their chosen answer. In line with self-determination theory, the quizzes and educational video clips, which were informed by BCT 21 and 22 (provide instruction on how to perform the behavior and model/demonstrate the behavior, respectively [65]), aimed at strengthening the individual competence of the young patients for managing their health condition [69].

The intervention component *coping planning tasks with social support (family member) moderated by a conversational agent* closed the experiential learning cycle. This component allowed patients to engage in abstract conceptualization [66] of the behavioral and cognitive skills that they had learned previously. In addition, this intervention component supported the improvement of asthma management as the intervention outcome. In line with self-determination theory [69], this further allowed patients to acquire overall increased competence via the integration of BCTs 7-9 (action planning, barrier identification/problem solving, and set graded tasks, respectively [65]).

Finally, we assume that there is a positive/negative reinforcement link that connects the outcomes of the conceptual model with the perceived characteristics of the app and working alliance with the conversational agent. This encourages patients to continue working with the conversational agent and increases engagement behavior, especially for young patients. That is, if neither the coaching sessions are perceived useful and joyful nor improvements of asthma management can be observed as a result of actual participation in the intervention, then engagement in the intervention will likely decrease, which has been shown in related interventions [81,82].

Technical Implementation

The intervention was developed with the open-source software platform MobileCoach [67,83,84], which has already been used successfully for various clinical and public health interventions [17,68,77-79,85,86] and ecological momentary assessments [87-89]. MobileCoach is available under the academia- and industry-friendly open-source Apache 2.0 license. MobileCoach-based interventions are delivered via SMS text messages, and via mobile apps for the Android and iOS operating systems. Moreover, MobileCoach-based interventions use a conversational agent for intervention delivery. MobileCoach client apps for iOS and Android use in-app encryption of user data, including password-protected access to MobileCoach Designer, a web-based interface for intervention authors, and a web-based MAX interface for health care professionals for chat interactions with human health coaches. Additionally, secure sockets layer (SSL) encoding was implemented to ensure privacy and safety of any data transfer between the mobile apps, the web-based MAX interface for health care professionals, MobileCoach Designer, and the MobileCoach server.

Costs of Intervention Components

Assessing the cost of each intervention component is relevant for real-life implementations [57]. Therefore, economic factors (eg, budgets of hospitals or health care professionals' time allocations) also need to be taken into consideration. The development costs of the MAX intervention, which is currently not classified as software related to a medical device in Switzerland as it remains a prototype and not a product, sums up to approximately US $250,000. This includes costs for the storybook, software development, project management, artwork, and production of video clips and the personalized QR code cards. Moreover, other costs per participant are linked to intervention components that involve either incentives (see above) or efforts by health care professionals. Regarding the latter, there are three such intervention components in the MAX intervention. First, health care professionals carry out the onboarding of patients, which encompasses two intervention components: (1) *demo of the asthma app by a trustful health care professional as a useful, easy to use, gamified, and autonomy-supporting tool*; and (2) *introduction of the conversational agent by a trustful health care professional as his/her digital personal assistant* (see Figure 1). Costs associated with these two intervention components are time needed for preparation, performance, and potential postprocessing of this task. Second, providers were involved in the assessment of short video clips sent by the participants as described by the intervention component *personalized feedback on inhalation technique by a health care professional*; see Figure 3 for a detailed explanation and illustration of this task and the according process. Here, associated costs concerned the time needed to assess the video clip and compile related feedback. This process was costly due to the economic costs of health care professionals but could be reduced to a certain extent by automatically providing a video tutorial showing how to perform the inhalation assessment with the web-based MAX interface for health care professionals. Since the video tutorial was integrated into every email that triggered an assessment, access to the tutorial was straightforward and thus of low cost. All other intervention components have low running costs as they are scalable due to their digital setup (eg, interaction over a mobile app and with a conversational agent, digital lottery, coaching sessions moderated by the conversational agent).

Evaluation of the Intervention

Study Design

MAX was assessed in a single-arm feasibility study in two home care settings offered by the Swiss Lung Association and in four secondary care settings at hospitals in the German-speaking part of Switzerland. The study was approved by the institutional review board of ETH Zurich (reference number: EK 2018-N-59).

Sample Size Considerations

The primary objective of this single-arm feasibility study was to develop, implement, and test the MAX intervention. Therefore, the study was exploratory by nature and did not include a detailed power analysis to determine a particular sample size. However, to identify a relevant amount of usability problems, at least 20 participants were required according to heuristics in usability engineering [90]. Moreover, to assess the potential reach of the intervention, we decided to approach between 90 and 100 participants.

Inclusion Criteria

The following inclusion criteria were defined and outlined in the corresponding intervention flyer (see Multimedia Appendix 12 and Multimedia Appendix 13)*:* (1) 10- to 15-year-old German-speaking patients diagnosed with asthma who have access to a smartphone with Google's Android (Version 4.1 or higher) or Apple's iOS (9.3 or higher) operating system and internet access via a data contract (3G/LTE) or wireless LAN (Wi-Fi) to watch the health literacy video clips, to interact with the conversational agent, and to fill out the online surveys; and (2) availability of a German-speaking family member of the patient (usually mother, father, or older sibling) who has access to a smartphone with internet capability via a data contract (3G/LTE) or wireless LAN (Wi-Fi) to be able to receive the SMS text messages from MAX and to fill out the online survey at the end of the intervention. This supporting family member must be motivated to support the young patient every second intervention day. There were no exclusion criteria.

Recruitment and Management of Study Participants

The participants were recruited during a 3-month period from January to April 2019 via participating health care professionals at six study sites in Switzerland. The study sites were two home care settings offered by the Swiss Lung Association (one in the canton Bern and one in the canton Thurgau) and four secondary care settings at hospitals in the German-speaking part of Switzerland. The health care professionals received instructions on how to install and use the mobile app before the start of the intervention. Additionally, health care professionals were provided with study instructions so they could consistently recruit and manage their patients during the study (see Multimedia Appendix 14 and Multimedia Appendix 15). This document (and every triggered email when an inhalation video clip was submitted) also included a link to a video tutorial that shows how to perform the inhalation assessment with the web-based MAX interface for health care professionals. They were also trained to introduce the MAX conversational agent as their personal digital assistant. The health care professionals recruited patients with a flyer that was personalized for each health care expert during their consultation hour (see Multimedia Appendix 12 and Multimedia Appendix 13 for examples of a personalized flyer), or via email, post, or telephone. Additionally, participants could access a website [91] for more information on the intervention (eg, with a demonstration video clip showing chatting with the MAX conversational agent), study participation, and frequently asked questions. If a candidate was interested in participating, inclusion criteria were checked by the health care professional, or, if the patient was not interested, corresponding reasons were noted down to better understand the patient's decisions (see Multimedia Appendix 10 and Multimedia Appendix 11). After reading more detailed study information (see Multimedia Appendix 16 for the German version) and signing the consent form (see Multimedia Appendix 17 for the German version), the health care professionals gave the patients their personal MAX intervention card in the form

of a business card with a QR code (see Multimedia Appendix 4). The QR code could be used with the standard photo app of a smartphone (capable of reading QR codes) and automatically forwarded the patient to either the Android or Apple store, depending on the type of their smartphone, to download the mobile app.

Measures

To assess the various aspects of the intervention, we used basic demographic, asthma-related, and intervention-related information, including age, gender, years since asthma diagnosis, the supporting family member during the intervention (eg, mother), mobile operating system used, and perceived uncertainty with asthma management. For the latter, this would include the conversation agent stating: "I have been taught some things about asthma by my development team, but I am still unsure from time to time. I'm sure you feel the same?" The answer options are "No, I am an asthma expert" (1), "I know quite well how to manage my asthma" (2), "Every now and then I feel insecure too" (3), and "Yes, I have been uncertain a lot before" (4). In addition, the following metrics and instruments were assessed.

The *reach* of the intervention was measured by the ratio of approached participants to those who started to interact with the conversational agent MAX on the mobile app. Reasons for nonparticipation were also gathered.

Working alliance between the patient and the conversational agent MAX was assessed with a German-adapted version of the Session Alliance Inventory [92] after coaching sessions 2, 8, and 14 (eg, "MAX and I respect each other" with answers anchored on a 7-point Likert scale ranging from 1 ["never"] to 8 ["always"]; see Multimedia Appendix 18).

Acceptance of the intervention was assessed in several ways. First, perceived usefulness ("The app helped me to increase my knowledge about my asthma"), ease of use ("The app was easy to use"), enjoyment ("I found the app enjoyable"), control ("I could control many aspects of the app"), and usage intention ("How much would you like to continue working with MAX?") were assessed by patients at the end of the intervention with instruments adapted from information systems research [50,93]. Single-item measures were used to reduce the burden of the intervention and answers were anchored on 7-point Likert scales ranging from "strongly disagree" (1) to "strongly agree" (7). Second, to obtain a more granular assessment for each coaching session, perceived usefulness ("Did you learn something new?" with answer options "No, I knew everything," "Yes, some new aspects," and "Yes, a lot of new aspects") and perceived enjoyment ("Did you enjoy today's lesson?" with answer options anchored on a 5-point Likert scale ranging from 1 ["strongly disagree"] to 5 ["strongly agree"]) were assessed at the end of each of the 14 coaching sessions randomly. A random assessment procedure with a maximum chance of 50% was implemented to reduce the burden of the intervention. If a participant had assessed the previous session, no assessment was triggered. Third, participation of the supporting family member ("Have you been supported today by the person you indicated?" with answer options "yes" and "no, unfortunately not") was measured at the end of each coaching session, which asked for social family support (ie, in sessions 1, 3, 5, 7, 8, 9, and 12). Fourth, during the setup procedure of the mobile app, we measured which of the two gender-specific characters of the MAX conversational agent (either the female *Maxime* or male *Maximilian*) was selected. Fifth, based on app usage data, we measured when participants dropped out of the intervention (ie, did not use it for 60 days). Sixth, we assessed the number of conversational turns between patients, health care professionals, and the conversational agent MAX. Finally, we also collected positive aspects of the intervention ("What did you really like about the intervention?") and suggestions for improvement ("What needs to be definitely changed in a future version?") from patients via an in-app conversation with the MAX conversational agent, the supporting family member via a web-based survey for which the MAX conversational agent sent a link by SMS to the family member, and health care professionals via a personal interview conducted by SH. All interview items are available in Multimedia Appendix 17.

Knowledge about asthma (ie, *cognitive skill*) was assessed at the beginning of the intervention (ie, at the end of the introductory chat with the conversational agent MAX) and in the last session by a validated health literacy quiz for children with asthma, with a quiz score ranging from 0 (no knowledge) to 11 (good knowledge) [80,94] (see Multimedia Appendix 19).

The inhalation technique of each patient (ie, *behavioral skill*) was systematically assessed by the patient's responsible health care professional with the help of predefined evaluation criteria These criteria were developed by health care professionals of the Swiss Lung Association and the participating pediatric pneumologists. The number of mistakes was counted, and it was decided for each assessment and health care professional whether there was a serious, potentially life-threatening, inhalation mistake.

Intervention completion rate was assessed by dividing the number of participants who finished the intervention within 60 days by the number of participants who started to interact with the conversational agent MAX.

We measured *human effort and responsiveness of health care professionals* to better understand the costs per patient related to the intervention. Here, these costs refer to (1) the onboarding time per patient, including a demo of the app and an introduction of the conversational agent MAX such as the time needed for preparation, performance, and potential postprocessing of this task; (2) the assessment of video clips with the time needed to assess the video clip and compile feedback; and (3) the number of conversational turns in the manual/human-managed chat channel of the MAX app. For the first cost aspect, we asked the health care professionals after the intervention to estimate the average onboarding time. For the second cost aspect, we objectively measured the duration required to review the video clips by health care professionals and the technical quality of the video clips (eg, "Did [patient name] exhale enough before inhalation?"). For the third cost aspect, we counted and compared the number of conversational turns between the patient and (a) the MAX conversational agent and (b) the health care professionals to better understand the extent to which the intervention can be delivered in a scalable way. In addition, we

measured the number of SMS reminders sent to patients and the supporting family members, since these also trigger costs.

Finally, we measured the time until patients received their feedback (ie, from the moment the video clip was submitted via the mobile app until the feedback was provided) as a further aspect of *human effort and responsiveness of health care professionals*.

Results

Participant Characteristics

The descriptive statistics of the study participants are shown in Table 1. Out of the 49 participants who started interacting with MAX, 33 were male with an average of 12 years of age and 5.6 years since receiving the asthma diagnosis. Only 13 of the participants indicated that they were uncertain a lot (n=2) or every now and then (n=11) in managing their asthma. The majority chose their mother as the supporting family member and iOS was used slightly more often than the Android operating system.

Table 1. Descriptive statistics of the patient-derived quantitative measures (N=49).

Construct	Respondents, n (%)	Mean (SD)
Demographic and asthma-related data		
Females	16 (33)	N/A[a]
Males	33 (67)	N/A
Age	49 (100)	12.04 (1.54)
Years since asthma diagnosis	39 (80)	5.61 (4.17)
Perceived uncertainty with asthma (measured in Coaching Session 4)	44 (90)	2.05 (0.81)
Mobile operating system		
Android	22 (45)	N/A
iOS	27 (55)	N/A
Supporting family member		
Mother	31 (63)	N/A
Father	9 (18)	N/A
Older brother	2 (4)	N/A
Older sister	3 (6)	N/A
Other	3 (6)	N/A
Patient-MAX CA[b] working alliance		
Coaching Session 2	44 (90)	6.34 (0.73)
Coaching Session 8	39 (80)	6.14 (0.96)
Coaching Session 14	36 (73)	6.34 (0.87)
Technology acceptance of mobile app		
Perceived usefulness	36 (73)	6.42 (1.09)
Perceived ease of use	36 (73)	6.75 (0.65)
Perceived enjoyment	36 (73)	6.47 (1.06)
Perceived control	36 (73)	5.53 (1.78)
Intention to continue working with the MAX CA	36 (73)	5.58 (1.73)
Perceived usefulness of coaching session		
Coaching Session 1	22 (45)	1.91 (0.68)
Coaching Session 2	10 (20)	2.50 (0.53)
Coaching Session 3	12 (24)	2.58 (0.67)
Coaching Session 4	14 (29)	2.36 (0.74)
Coaching Session 5	13 (27)	2.54 (0.78)
Coaching Session 6	14 (29)	2.29 (0.73)
Coaching Session 7	13 (27)	2.38 (0.77)
Coaching Session 8	13 (27)	2.31 (0.77)
Coaching Session 9	12 (24)	2.58 (0.67)
Coaching Session 10	16 (33)	2.50 (0.73)
Coaching Session 11	11 (22)	1.82 (0.75)
Coaching Session 12	13 (27)	2.38 (0.87)
Coaching Session 13	16 (33)	1.88 (0.62)
Coaching Session 14	13 (27)	2.15 (0.80)

Construct	Respondents, n (%)	Mean (SD)
Total	192 (100)	2.28 (0.74)
Perceived enjoyment of coaching session		
Coaching Session 1	22 (45)	4.91 (0.29)
Coaching Session 2	10 (20)	4.70 (0.48)
Coaching Session 3	12 (24)	4.83 (0.39)
Coaching Session 4	12 (24)	4.83 (0.39)
Coaching Session 5	13 (27)	4.69 (1.11)
Coaching Session 6	14 (29)	4.79 (0.58)
Coaching Session 7	13 (27)	5.00 (0.00)
Coaching Session 8	13 (27)	4.69 (0.48)
Coaching Session 9	12 (24)	4.83 (0.39)
Coaching Session 10	16 (33)	4.81 (0.54)
Coaching Session 11	11 (22)	4.64 (0.92)
Coaching Session 12	13 (27)	4.69 (0.85)
Coaching Session 13	16 (33)	4.75 (0.58)
Coaching Session 14	13 (27)	5.00 (0.00)
Total	190 (100)	4.81 (0.56)
Duration to complete the intervention/coaching session[c]		
Average duration (days)	37 (76)	21.46 (11.55)
Average days per coaching session	37 (76)	1.43 (0.77)
Conversational turns between the patients and the MAX CA		
Participants finishing the intervention	37 (76)	365.49 (11.85)
Participants not finishing the intervention	12 (25)	129.58 (59.86)
Conversational turns between the patients and health care professionals		
Participants finishing the intervention	37 (76)	1.68 (1.68)
Participants not finishing the intervention	12 (25)	1.00 (1.35)
In-app (free of cost) and SMS reminders sent to patients and supporting family member		
Participants finishing the intervention	37 (76)	11.57 (8.46)
Participants not finishing the intervention	12 (24)	20.75 (15.88)
SMS reminders sent to patients after 5 days of nonactivity		
Participants finishing the intervention	37 (76)	0.24 (0.86)
Participants not finishing the intervention	12 (24)	2.50 (1.68)
SMS reminders sent to supporting family member after 7 days of nonactivity		
Participants finishing the intervention	37 (76)	0.14 (0.67)
Participants not finishing the intervention	12 (24)	2.00 (1.28)
Asthma knowledge (cognitive skills)		
Asthma quiz score onboarding (pretest)	48 (98)	7.73 (2.24)
Asthma quiz score coaching session 14 (posttest)[d]	48 (98)	8.79 (2.27)

Construct	Respondents, n (%)	Mean (SD)
Asthma quiz score Coaching Session 14 (posttest)[e]	37 (76)	9.43 (1.76)

[a]N/A: not applicable.

[b]CA: conversational agent.

[c]Based on data from participants finishing the intervention.

[d]Last observation carried forward (ie, the pretest value was used for 11 participants).

[e]Complete cases, no missing values.

The flow chart of the MAX intervention, including details for nonparticipation and dropouts, is shown in Figure 4. Reach was 49.5% with 49 out of 99 approached patients downloading the app and starting to interact with the MAX conversational agent. Availability of a smartphone was the major reason for nonparticipation (n=14, 14%), and the most frequent dropouts happened during onboarding (n=3) and Coaching Session 6 (n=3). To better understand sessions after which participants no longer interacted with the MAX conversational agent (ie, they dropped out), Figure 5 indicates the key task involved in each "dropout session." The effort to complete a specific coaching session and disclosing personal information (eg, recording the inhalation technique with the face of the patient) may have led to dropout. Participants who finished the intervention (n=37) did so on average within 3 weeks, which was within the incentivized duration of 4 weeks.

Figure 4. Subject acquisition and participation flow chart. CA: conversational agent.

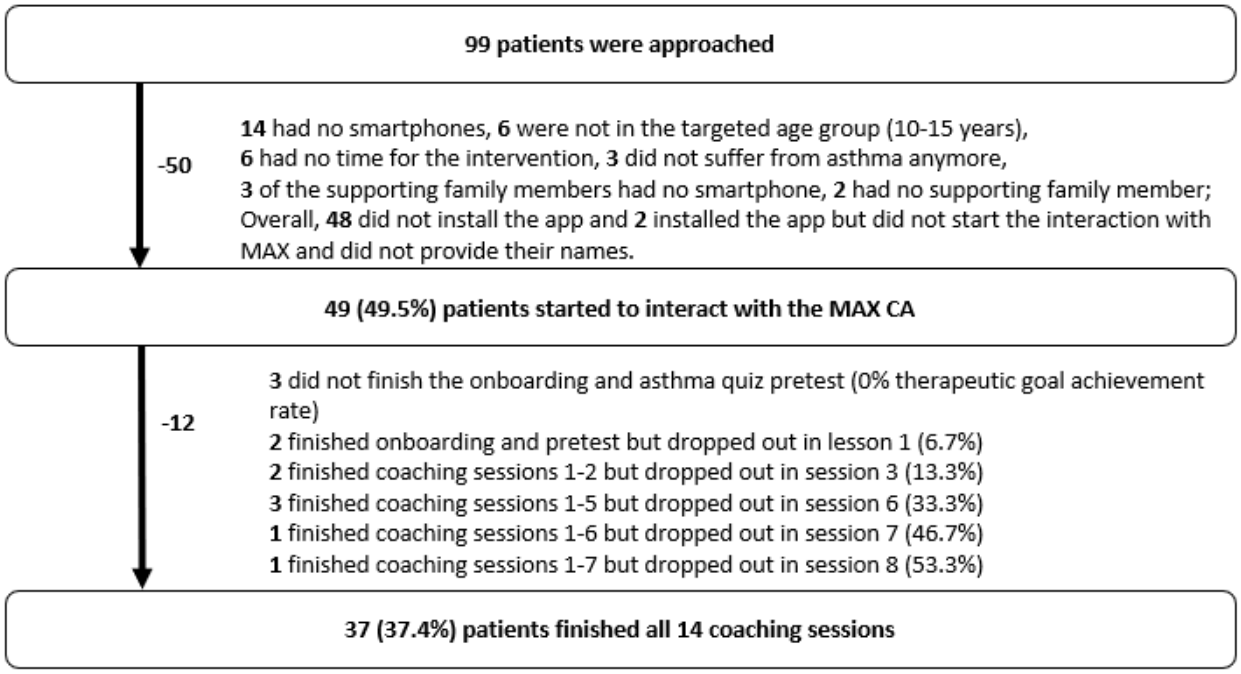
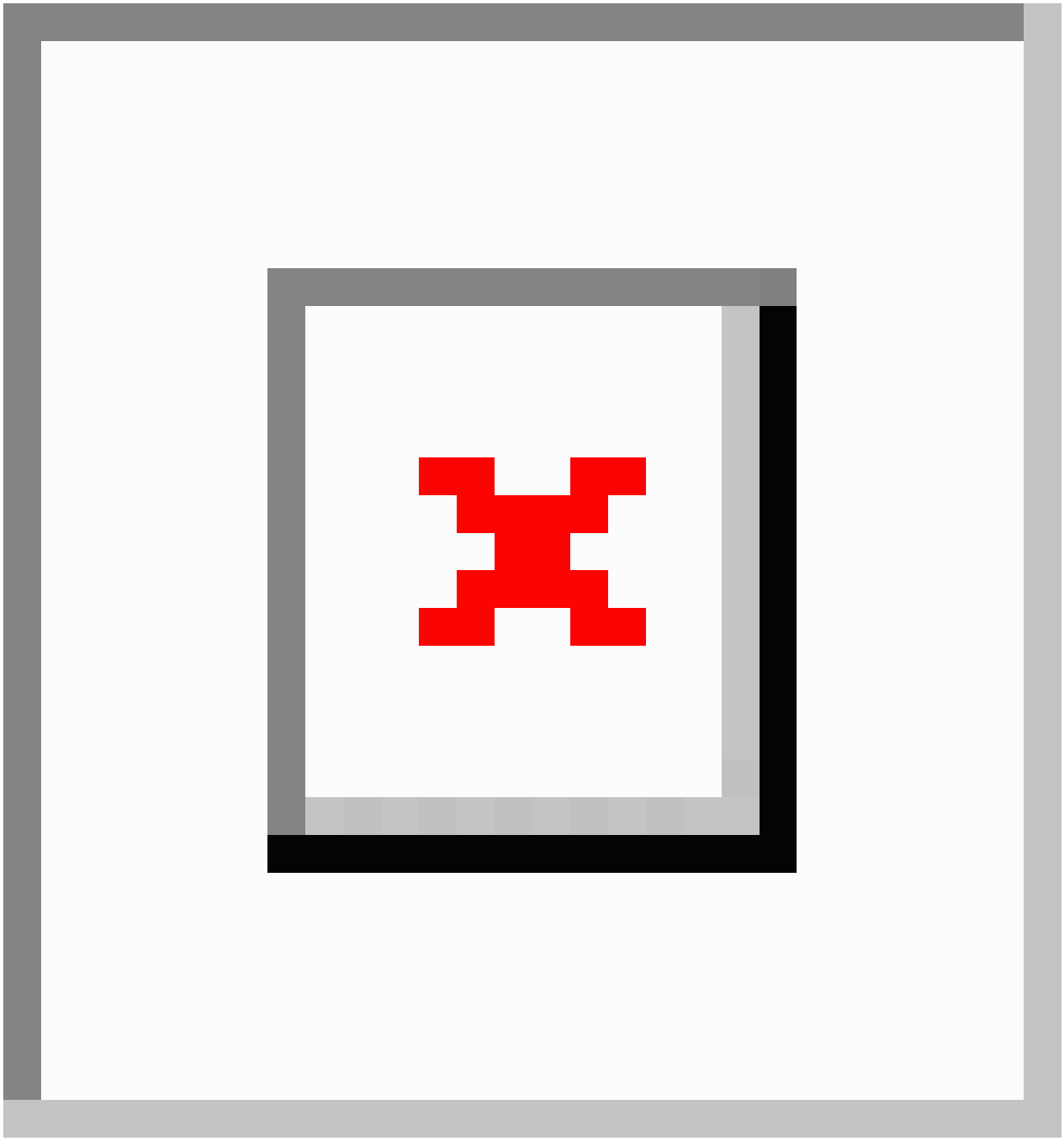


Figure 5. Participants for each coaching session and potential reasons for dropouts.



Session Alliance Inventory

The session alliance inventory indicated high working alliance ratings between the young patients and the MAX conversational agent from the very beginning of the intervention until the end (Table 1).

Technology Acceptance

Technology acceptance perceptions of the young patients regarding the mobile app are shown in Table 1. All mean values lie clearly above the neutral scale value of 4, indicating positive evaluations of the mobile app. Moreover, patients learned new aspects about asthma management and enjoyed the coaching sessions. Out of 275 coaching sessions, in which a family member was asked to support the young patients, patients indicated 269 times (97.8%) that they were supported by a family member. For the gender-specific choices of the MAX conversational agent, all participants chose the character corresponding to their own indicated gender.

Qualitative Feedback

The detailed qualitative feedback with representative quotes is provided in Multimedia Appendix 20. First, patients liked the educational content of the intervention and the text-based features of the conversational agent the most. Second, supporting family members also highlighted the educational content besides the experiential value of the intervention. Third, health care professionals positively emphasized the perceived ease of use and the significant supporting role of family members in this

intervention. With respect to improvement suggestions of the intervention, patients indicated that there was too much predefined text. This concern was also shared by supporting family members. Health care professionals indicated that lack of access to smartphones, especially for young patients, was a limiting factor to further increase the reach of the intervention. In addition, health care professionals indicated the following features to be considered in a future version. First, they would prefer an adaptation of the inclusion criteria, especially regarding the age range, to be able to further address younger and older patients. Second, they suggested cooperating with pneumologists and general practitioners to expand the intervention to other health-related topics or diseases (eg, eating disorders or diseases with similar complexity to asthma). Third, they suggested integrating further interaction between the health care professionals and patients (eg, follow-up questions).

Asthma Knowledge

Asthma knowledge (cognitive skills) scores at the beginning and end of the MAX intervention are shown in Table 1. Paired-sample *t* tests revealed a significant increase in scores and large effects with two approaches: a complete case analysis (n=37, t_{36}=–3.68, *P*<.001, *d*=1.19) and with the baseline observation carried forward (n=48, t_{47}=–3.54, *P*<.001, *d*=0.91).

Intervention Completion Rate

The intervention completion rate was 75.5%, as 37 out the 49 patients finished the intervention.

Overall, 42 inhalation video clips were recorded and submitted to the health care professionals (Table 2). All of these clips had sufficient technical quality for evaluation. The majority of inhalant medications used were dry powder inhaler and metered-dose inhaler. The health professionals' assessments of the inhalation techniques (behavioral skills) based on these video clips are listed in Table 3. In summary, health care professionals identified 0.9 inhalation mistakes in each video clip (N=42). For two video clips, three serious inhalation mistakes were identified, eliciting a feedback to resend a corrected video clip. After resubmission, no severe inhalation mistakes could be identified in the second video clip.

Table 2. Descriptive statistics of inhalation video clip assessments (N=42).

Variable	n (%)	Mean (SD)
Inhalant		
Dry powder inhaler (Turbuhaler)	17 (40)	N/A[a]
Metered-dose inhaler	16 (38)	N/A
Dry powder inhaler (Diskus)	9 (21)	N/A
Duration of video clip assessments (seconds)		
Primary care providers (N=2)	14 (33)	409.51 (346.48)
Secondary care providers (N=4)	28 (67)	126.94 (102.80)
Total	42 (100)	221.13 (251.39)
Inhalation mistakes identified per submitted video clip		
Primary care providers (N=2)	14 (33)	0.93 (0.83)
Secondary care providers (N=4)	28 (67)	0.93 (1.30)
Total	42 (100)	0.93 (1.16)
Days until feedback was provided (including weekends)		
Primary care providers (N=2)	14 (33)	2.25 (1.83)
Secondary care providers (N=4)	28 (67)	2.40 (1.81)
Total	42 (100)	2.34 (1.80)

[a]N/A: not applicable.

Table 3. Inhalation technique assessments by health care professionals.

Assessment question	Yes, n (%)	No, n (%)	Not visible on the video, n (%)
Questions for all assessments (N=42)			
Has _[a] the correct posture (ie, an upright upper body) during inhalation?	42 (100)	0 (0)	0 (0)
Did _ load/prepare the device correctly?	30 (71)	4 (10)	8 (19)
Did _ exhale enough before inhalation?	30 (71)	8 (19)	4 (10)
Did _ inhale deeply and long enough through the mouth during inhalation?	34 (81)	7 (17)	1 (2)
Did _ hold his breath for 5-10 seconds? OR an alternative for the metered-dose inhaler: Were 10 calm breaths taken via the upstream chamber?	38 (90)	4 (9.52%)	0 (0)
Did _ exhale slowly afterward?	33 (79)	5 (11.90%)	4 (10)
Additional metered-dose inhaler questions (N=16)			
Has the cap of the dosing aerosol been removed?	13 (81)	0 (0)	3 (19)
Was the metered dose aerosol shaken before inhalation?	10 (63)	2 (13)	4 (25)
Was the upstream chamber used?	15 (94)	1 (6)	0 (0)
Was the upstream chamber clean? (N=15[b])	14 (93)	0 (0)	1 (7)
Was the age-appropriate upstream chamber used? (mouthpiece, mask) (N=15[b])	15 (100)	0 (0)	0 (0)
Was there a whistling sound of the upstream chamber during inhalation? (inhaled too strongly and quickly)[c] (N=15[b])	11 (73)	3 (20)	1 (7)
Did _ trigger the device at the right time during inhalation?	15 (100)	0 (0)	0 (0)
Was exhaled incorrectly into the powder inhaler so that there is a risk of clumping?[c,d]	22 (85)	3 (12)	1 (4)
Has _ rinsed their mouth with water after inhalation or eaten anything?[e]	7 (22)	2 (6)	23 (72)

[a]_indicates the patient's name.

[b]Includes only those who answered "yes" to using the upstream chamber.

[c]Reverse coded.

[d]Additional question for those using the dry powder inhaler only (N=26).

[e]Additional question if the inhalant contained cortisol (N=32).

Human Effort and Responsiveness of Health Care Professionals

For the human effort and responsiveness of health care professionals (ie, to better understand the per-patient costs related to the intervention), the average time of the app onboarding process (excluding study-specific discussions) was approximately 15 minutes. Moreover, it took health care professionals an average of 221 seconds to assess the videos clips, with a clear difference observed between health care settings (average of 410 seconds in the primary care setting and 127 seconds in the secondary care setting; see Table 2). For the responsiveness of health care professionals, patients received feedback on their submitted video clips after an average of approximately 2.4 days (Table 2). In contrast to the assessment time, there were no differences between the health care professionals of the primary and secondary care settings. For the distribution of conversational turns, 99.5% (15,078/15,152) took place between patients and the MAX conversational agent, and only 0.5% (74/15,152) occurred between patients and health care professionals (Table 1). This indicates a very low amount of human effort (ie, 1-1.7 conversational turns between a health care professional and patient; see Table 1). Finally, between 0.1 and 2.5 SMS reminders were sent out on average per patient by the MAX conversational agent (Table 1), in addition to the 7 SMS text messages that were sent out to invite the supporting family members to join the seven "social support" coaching sessions.

The depersonalized data can be found on the Open Science Framework [95] for replication purposes and future analyses. It should be noted that not all data can be published due to ethical considerations and to protect the privacy of the participants of this study.

Discussion

Primary Findings

We have described the design of MAX, a smartphone-based and conversational agent–delivered asthma intervention that supports health care professionals targeting children-parent dyads in their everyday lives. Although there have been recent review papers discussing the use of conversational agents in health care [63,96-101], the current conversational agent is the first (to the best of our knowledge) that takes over the role of a scalable social actor framed as a scalable assistant of a health care professional that mediates communication among various relevant stakeholders in the context of chronic disease management. For this purpose, the MAX conversational agent uses several communication channels (eg, in-app chat, email, and SMS), and therefore "lives" not only on a smartphone in the pocket of a patient but is rather omnipresent (ie, MAX appears on the phones of family members, such as via SMS, or on desktop or tablet computers of health care professionals, such as via emails and the web-based MAX interface for health care professionals). This is also the first assessment of this type of mediating conversational agent outside of a lab setting, as many other conversational agents have been assessed [96], but rather in a realistic longitudinal intervention field study in a complex sociotechnical system with various stakeholders. With the MAX conversational agent, we were also able to show how to extend the reach of health care professionals into the everyday lives of patients in an efficient way without compromising the quality of care and working alliance. This is especially relevant in times of social distancing such as during the ongoing COVID-19 pandemic.

The design of MAX was driven by an interdisciplinary effort that resulted in a conceptual model with intervention components informed by human behavior and experiential learning theories [61,65,66], findings from technology acceptance research [49,51,58,102], and prior experiences of the authors with conversational agents that support health care professionals and young adolescent patients [67,68].

The results of this first feasibility study indicate an overall positive evaluation with respect to the reach of the intervention (ie, 49.5% of 99 young patients approached did install the app and started to interact with the MAX conversational agent), the strong working alliance between patients and the MAX conversational agent, and high acceptance of the intervention by all relevant stakeholders (ie, health care professionals, young patients, and their supporting family members). Compared to very similar conversational agent research targeting childhood obesity [68], physical inactivity [77], or the management of chronic pain [17], this intervention resulted in a high overall therapeutic goal achievement rate (75.5%) but also in improved asthma knowledge test scores and behavioral skills (ie, no identified inhalation mistakes occurred after the feedback from health care professionals). Moreover, the MAX conversational agent was able to motivate family members to support the young patients most of the time when asked (97.8%). In terms of human effort and responsiveness of health care professionals, it can be concluded that the MAX intervention is scalable since most of the conversational turns (99.5%) involved the patients and the MAX conversational agent. After the app onboarding process, which takes an average of 15 minutes, health care professionals had, on average, only one conversational turn with the patients via the manual chat channel of the MAX app when they provided their personalized feedback regarding the inhalation technique. In addition, it took them less than 4 minutes to assess the inhalation technique and 3 days to deliver that feedback to the patients. For each patient, this intervention involved an average of 20 minutes of human effort, 10 automated SMS text messages, including three reminders, and additional costs for gift vouchers, including lottery winnings. We minimized the risk for smartphone addiction [76] by limiting the amount of possible sessions to one per day and further including active exercises outside the digital environment of the app to increase social interaction and to counteract increasing smartphone usage among children [103].

The qualitative feedback suggested several valued and important features, as well as challenges and potential improvements of the intervention. Combining results from each question of the quantitative analysis, and considering the importance and frequency mentioned, several aspects must be discussed and eventually improved in future versions. First, technical issues should be limited as the reach and effectiveness of such an intervention is dependent on problem-free operation. This requires, based on the experience gathered with the MAX intervention, a better understanding and analysis of the technical infrastructure of the health care professionals' institutions (eg, simple-to-use patient access to broadband internet via Wi-Fi in hospitals). Even though the text-based conversational agent was perceived as positive and engaging, participants indicated that the conversational agent had too many predefined answer options. It was previously suggested that conversational agents can be influential and engaging for young patients, and that open-text answers are highly appreciated [104]. However, privacy issues with conversational agents and open-text answers were pointed out by prior work [105], as conversational agents that are responsive to such inputs could potentially and unintentionally retrieve more and more personal information.

Limitations and Future Work

This study was designed as a feasibility study with a limited number of participants. It therefore provides the basis, and not the end solution, for future activities in the field. Based on our limited sample, it is clear that the results are not representative and must be interpreted with caution. Further, only health care professionals from four hospitals (eg, pediatric pneumologists) and two cantonal patient organizations of the Swiss Lung Association participated in this study. Therefore, it is not clear whether and to what degree the MAX intervention would work the same way with other relevant health care professionals such as a general practitioner. These nonspecialized health care experts may require significantly more time for the assessment of the inhalation video clips or would not have the expertise to do so without additional educational efforts. Another limitation of this study pertains to the inductive open coding of the interviews that was performed by one author only (SH), resulting in a certain bias of the qualitative results. In addition, since the social support assessment was self-reported by the young

patients and linked to additional points for the MAX intervention (to increase the chance to be among the winners), it can be assumed that the supportive involvement of family members was overestimated. Furthermore, the web-based MAX interface for health care professionals (and with it the patient data) was not integrated into hospital information systems or the information system of the patient organization. Specifically, some data had to be stored in a redundant way (eg, contact number, patient name) in the MAX system, which likely resulted in an overestimation of efforts (eg, the duration of the onboarding process). Finally, we have only reported the costs and efforts related to the MAX intervention, and therefore no implications regarding cost-effectiveness can be discussed. It is therefore important that future work investigates the costs of asthma management (eg, the frequency and costs of hospitalizations due to asthma exacerbations) and to which degree they can be reduced with the MAX intervention.

The MAX intervention itself can be improved in several ways. First and foremost, as a next step according to the multiphase optimization strategy [57], we suggest performing optimization trials to identify intervention components that have a positive and significant impact on cognitive and behavioral skills. Toward this end, we suggest assessing the components that are more costly (ie, intervention components that involve human effort). The resulting "effective" intervention package should then be assessed in a final randomized controlled trial with relevant distal health outcomes such as asthma control or quality of life. Moreover, we suggest incorporating a digital biomarker that is able to predict life-threatening events (eg, asthma attacks). For example, there is evidence that the number of nocturnal cough events is negatively correlated with asthma control [106,107], and that nocturnal cough in adult asthma patients can be detected reliably with the microphone of a smartphone [89,108]. Having such a digital biomarker may also help to further develop the MAX intervention as a just-in-time adaptive intervention (JITAI) [109,110]. In such an intervention, after the basic psychoeducational coaching sessions are finished, the MAX conversational agent would message patients only when a specific state of vulnerability [89] and state of receptivity [111] are identified. In addition, and consistent with the JITAI approach, one may also consider an intervention component that monitors medication intake and sends out medication reminders in case no inhalation events were detected. The systematic assessment of inhalation video clips by health care professionals can also be used as a label for the correct use of inhalation devices. Additionally, taking advantage of these labels and the latest advances in video classification methods for activity detection [112] may enable the automatic assessment of inhalation technology. As a consequence, this may reduce the time required to assess the inhalation technique, and may even increase the quality of the assessments. Furthermore, since there was a clear difference in the assessment time of the inhalation video clips between the primary and secondary care settings, a dedicated and specialized expert may be considered for this task. However, this addition may undermine the working alliance between the patient and the primary point of contact (ie, the health care professional who takes care of that patient). Finally, future deployments of MAX must consider a robust, technical infrastructure with a clear focus on the easiest Wi-Fi access possible during on-site consultations to guarantee an efficient download of the app and onboarding process.

In case none of these additional intervention components or studies is considered, estimates of the MAX project team indicate that the development of the current MAX intervention into a "product" would cost another US $100,000. General ongoing costs include keeping the intervention content updated according to recent asthma management guidelines (at a cost of approximately US $10,000 every 3 years) and maintaining technical software (at a cost of approximately US $10,000 per year). In addition, an appropriate legal framework and incentive mechanism has to be established in the Swiss health care system that allows prescription of this "digital pill" by health care providers so that the human and technical efforts, as well as the incentives for the participants (eg, rewards based on the number of intervention points they achieve), can be covered by corresponding payers (eg, individuals or health insurance companies). The recently implemented Digital Healthcare Act in Germany can serve as an example in this regard.

Comparison With Prior Work

Digital health interventions for asthma include numerous mobile health apps that provide patients with information and help them track symptoms or medications, often using a gamification component [45-47]. A systematic review of 15 different digital interventions for pediatric asthma management showed that 87% of the interventions improved medication and behavioral adherence, and 53% demonstrated improved health outcomes [113]. Although these mobile health apps offer a range of features (eg, automated personalized texts, interactive websites, and online modules) to inform patients about asthma, they have not included scalable text-based health care conversational agents to support communication with health care professionals. Previous studies in other health domains have demonstrated promising results in using conversational agents to improve outcomes, such as promoting physical activity for childhood obesity [104,114]. By applying a scalable conversational agent for asthma specifically, the MAX intervention can provide greater health care professional interaction at reduced cost, which has been a key concern in past asthma interventions [113]. A unique advantage of MAX is its use of a three-component intervention that involves health care professionals, the digital assistant MAX, and family members to support young patients as they work on specific tasks to expand asthma knowledge and improve behavioral skills.

Conclusions

We have shown that conversational agents framed as digital assistants of health care professionals have the potential to improve cognitive and behavioral skills in chronic disease management, with asthma in children as one specific example. We have demonstrated that conversational agents can take over the role of a mediating social actor in a complex health care setting with various stakeholders, and deliver a digital health intervention in a scalable way into the everyday life of patients and their family members. Consistent with the novel JITAI approach, this study provides further insights into the use of conversational agents that, in the future, may "listen into" states of vulnerability and states of receptivity and, as a result, direct

relevant information to appropriate individuals, including the patient, a romantic partner, family member, a nurse, or medical doctor. We therefore envision a future in which scalable conversational agents act like a grand maestro, who dynamically directs an orchestra through a symphony of life based on what the various musicians offer and he or she perceives, and, with each repetition, gets better and better in doing so.

Conflicts of Interest

TK, TS, SH, FB, EF, and FW are affiliated with the Centre for Digital Health Interventions (CDHI), a joint initiative of the Department of Management, Technology and Economics at ETH Zurich, and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. The MAX intervention and study were cofunded by CSS and the Swiss Lung Association. TK is also cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways and has used the open source MobileCoach platform for that purpose. Pathmate Technologies received funding from the Swiss Lung Association to develop the MAX app and the web-based MAX interface for health care professionals based on the MobileCoach software. The developed generic software modules were made open source in the latest version of MobileCoach by the CDHI. Neither CSS nor the Swiss Lung Association or Pathmate Technologies was involved in any aspect of data analysis or manuscript preparation. None of the health care professionals was involved in the data analysis in any aspect. Neither CSS nor Pathmate Technologies was involved in the study design. All other authors report no conflicts of interest.

Multimedia Appendix 1
Overview of intervention coaching sessions and schedule (long).
[PDF File (Adobe PDF File), 1279 KB - jmir_v23i2e25060_app1.pdf]

Multimedia Appendix 2
Overview of intervention coaching sessions and schedule (short).
[PDF File (Adobe PDF File), 1227 KB - jmir_v23i2e25060_app2.pdf]

Multimedia Appendix 3
Web-based MAX interface for health care professionals.
[PDF File (Adobe PDF File), 1138 KB - jmir_v23i2e25060_app3.pdf]

Multimedia Appendix 4
Physical onboarding card.
[PDF File (Adobe PDF File), 909 KB - jmir_v23i2e25060_app4.pdf]

Multimedia Appendix 5
Screenshots of the MAX app.
[PDF File (Adobe PDF File), 1051 KB - jmir_v23i2e25060_app5.pdf]

Multimedia Appendix 6
Video Onboarding, quiz, in-app video.
[MP4 File (MP4 Video), 1729 KB - jmir_v23i2e25060_app6.mp4]

Multimedia Appendix 7
Video clip of patient inhalation (German).
[MP4 File (MP4 Video), 14678 KB - jmir_v23i2e25060_app7.mp4]

Multimedia Appendix 8
Screenplay MobileCoach Asthma (German).
[PDF File (Adobe PDF File), 2306 KB - jmir_v23i2e25060_app8.pdf]

Multimedia Appendix 9
Screenplay MobileCoach Asthma (English).
[PDF File (Adobe PDF File), 2155 KB - jmir_v23i2e25060_app9.pdf]

Multimedia Appendix 10

Study recruitment assessment (German).
[PDF File (Adobe PDF File), 743 KB - jmir_v23i2e25060_app10.pdf]

Multimedia Appendix 11
Study recruitment assessment (English).
[PDF File (Adobe PDF File), 814 KB - jmir_v23i2e25060_app11.pdf]

Multimedia Appendix 12
Original study flyer (German).
[PDF File (Adobe PDF File), 1610 KB - jmir_v23i2e25060_app12.pdf]

Multimedia Appendix 13
Study flyer (English).
[PDF File (Adobe PDF File), 817 KB - jmir_v23i2e25060_app13.pdf]

Multimedia Appendix 14
Study information for health care professionals (German).
[PDF File (Adobe PDF File), 840 KB - jmir_v23i2e25060_app14.pdf]

Multimedia Appendix 15
Video clip: explanation of inhalation assessment.
[MP4 File (MP4 Video), 17627 KB - jmir_v23i2e25060_app15.mp4]

Multimedia Appendix 16
Study information for patients and family members (German).
[PDF File (Adobe PDF File), 757 KB - jmir_v23i2e25060_app16.pdf]

Multimedia Appendix 17
Study consent for patients and family members (German).
[PDF File (Adobe PDF File), 643 KB - jmir_v23i2e25060_app17.pdf]

Multimedia Appendix 18
Survey instruments.
[PDF File (Adobe PDF File), 1815 KB - jmir_v23i2e25060_app18.pdf]

Multimedia Appendix 19
Health Literacy quiz items (English).
[PDF File (Adobe PDF File), 111 KB - jmir_v23i2e25060_app19.pdf]

Multimedia Appendix 20
Qualitative feedback.
[PDF File (Adobe PDF File), 3402 KB - jmir_v23i2e25060_app20.pdf]

References

1. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol 2016 Mar;34(3):239-246. [doi: 10.1038/nbt.3495] [Medline: 26963544]
2. Frey U, Suki B. Complexity of chronic asthma and chronic obstructive pulmonary disease: implications for risk assessment, and disease progression and control. Lancet 2008 Sep 20;372(9643):1088-1099 [FREE Full text] [doi: 10.1016/S0140-6736(08)61450-6] [Medline: 18805337]
3. Dimatteo MR. The role of effective communication with children and their families in fostering adherence to pediatric regimens. Patient Educ Couns 2004 Dec;55(3):339-344. [doi: 10.1016/j.pec.2003.04.003] [Medline: 15582339]
4. Mackey LM, Doody C, Werner EL, Fullen B. Self-management skills in chronic disease management: what role does health literacy have? Med Decis Making 2016 Aug;36(6):741-759. [doi: 10.1177/0272989X16638330] [Medline: 27053527]
5. Hill-Briggs F. Problem solving in diabetes self-management: a model of chronic illness self-management behavior. Ann Behav Med 2003;25(3):182-193. [doi: 10.1207/S15324796ABM2503_04] [Medline: 12763713]

6. Barlow JH, Ellard DR. Psycho-educational interventions for children with chronic disease, parents and siblings: an overview of the research evidence base. Child Care Health Dev 2004 Nov;30(6):637-645. [doi: 10.1111/j.1365-2214.2004.00474.x] [Medline: 15527474]
7. Voruganti T, Grunfeld E, Makuwaza T, Bender JL. Web-based tools for text-based patient-provider communication in chronic conditions: scoping review. J Med Internet Res 2017 Oct 27;19(10):e366 [FREE Full text] [doi: 10.2196/jmir.7987] [Medline: 29079552]
8. Alpert JM, Dyer KE, Lafata JE. Patient-centered communication in digital medical encounters. Patient Educ Couns 2017 Oct;100(10):1852-1858 [FREE Full text] [doi: 10.1016/j.pec.2017.04.019] [Medline: 28522229]
9. Brigden A, Anderson E, Linney C, Morris R, Parslow R, Serafimova T, et al. Digital behavior change interventions for younger children with chronic health conditions: systematic review. J Med Internet Res 2020 Jul 31;22(7):e16924 [FREE Full text] [doi: 10.2196/16924] [Medline: 32735227]
10. Stukus DR, Farooqui N, Strothman K, Ryan K, Zhao S, Stevens JH, et al. Real-world evaluation of a mobile health application in children with asthma. Ann Allergy Asthma Immunol 2018 Apr;120(4):395-400. [doi: 10.1016/j.anai.2018.02.006] [Medline: 29452259]
11. Mosnaim G, Li H, Martin M, Richardson D, Belice PJ, Avery E, et al. A tailored mobile health intervention to improve adherence and asthma control in minority adolescents. J Allergy Clin Immunol Pract 2015;3(2):288-290 [FREE Full text] [doi: 10.1016/j.jaip.2014.10.011] [Medline: 25609351]
12. Morton RW, Elphick HE, Rigby AS, Daw WJ, King DA, Smith LJ, et al. STAAR: a randomised controlled trial of electronic adherence monitoring with reminder alarms and feedback to improve clinical outcomes for children with asthma. Thorax 2017 Apr;72(4):347-354. [doi: 10.1136/thoraxjnl-2015-208171] [Medline: 27815524]
13. Kenyon CC, Chang J, Wynter S, Fowler JC, Long J, Bryant-Stephens TC. Electronic adherence monitoring in a high-utilizing pediatric asthma cohort: a feasibility study. JMIR Res Protoc 2016 Jun 22;5(2):e132 [FREE Full text] [doi: 10.2196/resprot.5362] [Medline: 27335355]
14. Chan AHY, Stewart AW, Harrison J, Camargo CA, Black PN, Mitchell EA. The effect of an electronic monitoring device with audiovisual reminder function on adherence to inhaled corticosteroids and school attendance in children with asthma: a randomised controlled trial. Lancet Respir Med 2015 Mar;3(3):210-219. [doi: 10.1016/S2213-2600(15)00008-9] [Medline: 25617215]
15. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 2010 Jul 01;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
16. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput Hum Interact 2005 Jun;12(2):293-327. [doi: 10.1145/1067860.1067867]
17. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 03;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
18. Ma T, Sharifi H, Chattopadhyay D. Virtual humans in health-related interventions: a meta-analysis. 2019 Presented at: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems; 2019; Glasgow p. 1-6. [doi: 10.1145/3290607.3312853]
19. Bickmore TW, Paasche-Orlow MK. The role of information technology in health literacy research. J Health Commun 2012;17(Suppl 3):23-29. [doi: 10.1080/10810730.2012.712626] [Medline: 23030559]
20. Asthma Fact Sheet. World Health Organization. URL: https://www.who.int/en/news-room/fact-sheets/detail/asthma [accessed 2020-04-13]
21. Patadia MO, Murrill LL, Corey J. Asthma: symptoms and presentation. Otolaryngol Clin North Am 2014 Feb;47(1):23-32. [doi: 10.1016/j.otc.2013.10.001] [Medline: 24286676]
22. Barrios RJ, Kheradmand F, Batts L, Corry DB. Asthma: pathology and pathophysiology. Arch Pathol Lab Med 2006 Apr;130(4):447-451 [FREE Full text] [doi: 10.1043/1543-2165(2006)130[447:APAP]2.0.CO;2] [Medline: 16594736]
23. Nunes C, Pereira AM, Morais-Almeida M. Asthma costs and social impact. Asthma Res Pract 2017;3:1-11 [FREE Full text] [doi: 10.1186/s40733-016-0029-3] [Medline: 28078100]
24. Eder W, Ege MJ, von Mutius E. The asthma epidemic. N Engl J Med 2006 Nov 23;355(21):2226-2235. [doi: 10.1056/NEJMra054308] [Medline: 17124020]
25. Haughney J, Price D, Kaplan A, Chrystyn H, Horne R, May N, et al. Achieving asthma control in practice: understanding the reasons for poor control. Respir Med 2008 Dec;102(12):1681-1693 [FREE Full text] [doi: 10.1016/j.rmed.2008.08.003] [Medline: 18815019]
26. Dewalt DA, Berkman ND, Sheridan S, Lohr KN, Pignone MP. Literacy and health outcomes: a systematic review of the literature. J Gen Intern Med 2004 Dec;19(12):1228-1239 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.40153.x] [Medline: 15610334]
27. Wang K, Chu N, Lin S, Chiang I, Perng W, Lai H. Examining the causal model linking health literacy to health outcomes of asthma patients. J Clin Nurs 2014 Jul;23(13-14):2031-2042. [doi: 10.1111/jocn.12434] [Medline: 24329740]

28. Lindquist LA, Go L, Fleisher J, Jain N, Friesema E, Baker DW. Relationship of health literacy to intentional and unintentional non-adherence of hospital discharge medications. J Gen Intern Med 2012 Feb;27(2):173-178 [FREE Full text] [doi: 10.1007/s11606-011-1886-3] [Medline: 21971600]
29. Jaffee EG, Arora VM, Matthiesen MI, Meltzer DO, Press VG. Health literacy and hospital length of stay: an inpatient cohort study. J Hosp Med 2017 Dec;12(12):969-973. [doi: 10.12788/jhm.2848] [Medline: 29236095]
30. Miles C, Arden-Close E, Thomas M, Bruton A, Yardley L, Hankins M, et al. Barriers and facilitators of effective self-management in asthma: systematic review and thematic synthesis of patient and healthcare professional views. NPJ Prim Care Respir Med 2017 Oct 09;27(1):57. [doi: 10.1038/s41533-017-0056-4] [Medline: 28993623]
31. Chang C. Asthma in children and adolescents: a comprehensive approach to diagnosis and management. Clin Rev Allergy Immunol 2012 Aug;43(1-2):98-137 [FREE Full text] [doi: 10.1007/s12016-011-8261-3] [Medline: 22187333]
32. Walia M, Paul L, Satyavani A, Lodha R, Kalaivani M, Kabra SK. Assessment of inhalation technique and determinants of incorrect performance among children with asthma. Pediatr Pulmonol 2006 Nov;41(11):1082-1087. [doi: 10.1002/ppul.20498] [Medline: 16989001]
33. Wade SL, Holden G, Lynn H, Mitchell H, Ewart C. Cognitive-behavioral predictors of asthma morbidity in inner-city children. J Dev Behav Pediatr 2000 Oct;21(5):340-346. [doi: 10.1097/00004703-200010000-00004] [Medline: 11064961]
34. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
35. Blake KV. Improving adherence to asthma medications: current knowledge and future perspectives. Curr Opin Pulm Med 2017 Jan;23(1):62-70. [doi: 10.1097/MCP.0000000000000334] [Medline: 27755160]
36. Searle A, Jago R, Henderson J, Turner KM. Children's, parents' and health professionals' views on the management of childhood asthma: a qualitative study. NPJ Prim Care Respir Med 2017 Sep 11;27(1):53. [doi: 10.1038/s41533-017-0053-7] [Medline: 28894094]
37. Kuprys-Lipinska I, Kolacinska-Flont M, Kuna P. New approach to intermittent and mild asthma therapy: evolution or revolution in the GINA guidelines? Clin Transl Allergy 2020;10:19 [FREE Full text] [doi: 10.1186/s13601-020-00316-z] [Medline: 32514334]
38. Potter PC. Current guidelines for the management of asthma in young children. Allergy Asthma Immunol Res 2010 Jan;2(1):1-13 [FREE Full text] [doi: 10.4168/aair.2010.2.1.1] [Medline: 20224672]
39. Brand PLP. Key issues in inhalation therapy in children. Curr Med Res Opin 2005;21(Suppl 4):S27-S32. [doi: 10.1185/030079905X61767] [Medline: 16138942]
40. de Benedictis D, Bush A. Asthma in adolescence: Is there any news? Pediatr Pulmonol 2017 Jan;52(1):129-138. [doi: 10.1002/ppul.23498] [Medline: 27273858]
41. Melani AS, Bonavia M, Cilenti V, Cinti C, Lodi M, Martucci P, Gruppo Educazionale Associazione Italiana Pneumologi Ospedalieri. Inhaler mishandling remains common in real life and is associated with reduced disease control. Respir Med 2011 Jun;105(6):930-938 [FREE Full text] [doi: 10.1016/j.rmed.2011.01.005] [Medline: 21367593]
42. Giraud V, Roche N. Misuse of corticosteroid metered-dose inhaler is associated with decreased asthma stability. Eur Respir J 2002 Feb;19(2):246-251 [FREE Full text] [doi: 10.1183/09031936.02.00218402] [Medline: 11866004]
43. Lavorini F, Magnan A, Dubus JC, Voshaar T, Corbetta L, Broeders M, et al. Effect of incorrect use of dry powder inhalers on management of patients with asthma and COPD. Respir Med 2008 Apr;102(4):593-604 [FREE Full text] [doi: 10.1016/j.rmed.2007.11.003] [Medline: 18083019]
44. Harris K, Mosler G, Grigg J. Theory-based self-management intervention to improve adolescents' asthma control: a cluster randomised controlled trial protocol. BMJ Open 2019 Apr 23;9(4):e025867 [FREE Full text] [doi: 10.1136/bmjopen-2018-025867] [Medline: 31015270]
45. Huckvale K, Morrison C, Ouyang J, Ghaghda A, Car J. The evolution of mobile apps for asthma: an updated systematic assessment of content and tools. BMC Med 2015 Mar 23;13:58 [FREE Full text] [doi: 10.1186/s12916-015-0303-x] [Medline: 25857569]
46. Grossman B, Conner S, Mosnaim G, Albers J, Leigh J, Jones S, et al. Application of human augmentics: a persuasive asthma inhaler. J Biomed Inform 2017 Mar;67:51-58 [FREE Full text] [doi: 10.1016/j.jbi.2017.02.003] [Medline: 28193465]
47. Tinschert P, Jakob R, Barata F, Kramer J, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. JMIR Mhealth Uhealth 2017 Aug 02;5(8):e113 [FREE Full text] [doi: 10.2196/mhealth.7177] [Medline: 28768606]
48. King P, Tester J. The landscape of persuasive technologies. Commun ACM 1999 May;42(5):31-38. [doi: 10.1145/301353.301398]
49. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quart 2012;36(1):157-178. [doi: 10.2307/41410412]
50. van der Heijden H. User acceptance of hedonic information systems. MIS Quart 2004;28(4):695-704. [doi: 10.2307/25148660]
51. Liu D, Santhanam R, Webster J. Toward meaningful engagement: a framework for design and research of gamified information systems. MIS Quart 2017 Apr 4;41(4):1011-1034. [doi: 10.25300/misq/2017/41.4.01]

52. Sanders C, Rogers A, Bowen R, Bower P, Hirani S, Cartwright M, et al. Exploring barriers to participation and adoption of telehealth and telecare within the Whole System Demonstrator trial: a qualitative study. BMC Health Serv Res 2012 Jul 26;12:220 [FREE Full text] [doi: 10.1186/1472-6963-12-220] [Medline: 22834978]
53. Sánchez-Morillo D, Crespo M, León A, Crespo Foix LF. A novel multimodal tool for telemonitoring patients with COPD. Inform Health Soc Care 2015 Jan;40(1):1-22. [doi: 10.3109/17538157.2013.872114] [Medline: 24380372]
54. Sanchez-Morillo D, Fernandez-Granero MA, Leon-Jimenez A. Use of predictive algorithms in-home monitoring of chronic obstructive pulmonary disease and asthma: A systematic review. Chron Respir Dis 2016 Aug;13(3):264-283 [FREE Full text] [doi: 10.1177/1479972316642365] [Medline: 27097638]
55. Klok T, Kaptein AA, Brand PLP. Non-adherence in children with asthma reviewed: The need for improvement of asthma care and medical education. Pediatr Allergy Immunol 2015 May;26(3):197-205. [doi: 10.1111/pai.12362] [Medline: 25704083]
56. Keim-Malpass J, Letzkus LC, Kennedy C. Parent/caregiver health literacy among children with special health care needs: a systematic review of the literature. BMC Pediatr 2015 Aug 05;15(1):92 [FREE Full text] [doi: 10.1186/s12887-015-0412-x] [Medline: 26242306]
57. Collins LM. Optimization of Behavioral, Biobehavioral, and Biomedical Interventions. USA: Springer; 2018.
58. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quart 2003;27(3):425-478. [doi: 10.2307/30036540]
59. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. J Counsel Psychol 1989 Apr;36(2):223-233. [doi: 10.1037/0022-0167.36.2.223]
60. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy (Chic) 2018 Dec;55(4):316-340. [doi: 10.1037/pst0000172] [Medline: 29792475]
61. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
62. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied conversational agents in clinical psychology: a scoping review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
63. ter Stal S, Kramer LL, Tabak M, op den Akker H, Hermens H. Design features of embodied conversational agents in eHealth: a literature review. Int J Hum Comput Stud 2020 Jun;138:102409. [doi: 10.1016/j.ijhcs.2020.102409]
64. Cassell J, Sullivan J, Prevost S, Churchill E. Embodied Conversational Agents. USA: MIT Press; 2000.
65. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
66. Kolb DA. Experience as the source of learning and development. Upper Sadle River: Prentice Hall; 1984.
67. Kowatsch T, Volland D, Shih I, Rüegger D, Künzler F, Barata F. Design and evaluation of a mobile chat app for the open source behavioral health intervention platform MobileCoach. 2017 Presented at: International Conference on Design Science Research Information Systems Technology; 2017; Germany p. 485-489. [doi: 10.1007/978-3-319-59144-5_36]
68. Kowatsch T, Nißen M, Shih I, Rüegger D, Volland D, Filler A, et al. Text-based healthcare chatbots supporting patient and health professional teams: preliminary results of a randomized controlled trial on childhood obesity. 2017 Presented at: Persuasive Embodied Agents for Behavior Change (PEACH2017) Workshop, co-located with the 17th International Conference on Intelligent Virtual Agents (IVA 2017); August 27, 2017; Stockholm, Sweden. [doi: 10.3929/ethz-b-000218776]
69. Ryan RM, Deci EL. Self-determination theory. Basic psychological needs in motivation, development and wellness. New York, NY: Guilford Press; 2018.
70. Deci EL, Olafsen AH, Ryan RM. Self-determination theory in work organizations: the state of a science. Annu Rev Organ Psychol Organ Behav 2017 Mar 21;4(1):19-43. [doi: 10.1146/annurev-orgpsych-032516-113108]
71. Ajzen I. The theory of planned behavior. Organization Behav Hum Decis Proc 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
72. Thom DH, Hall MA, Pawlson LG. Measuring patients' trust in physicians when assessing quality of care. Health Aff (Millwood) 2004;23(4):124-132. [doi: 10.1377/hlthaff.23.4.124] [Medline: 15318572]
73. Birkhäuer J, Gaab J, Kossowsky J, Hasler S, Krummenacher P, Werner C, et al. Trust in the health care professional and health outcome: A meta-analysis. PLoS One 2017;12(2):e0170988 [FREE Full text] [doi: 10.1371/journal.pone.0170988] [Medline: 28170443]
74. Aron A, Aron EN, Smollan D. Inclusion of other in the self scale and the structure of interpersonal closeness. J Pers Soc Psychol 1992;63(4):596-612. [doi: 10.1037/0022-3514.63.4.596]
75. Nass C, Steuer J, Tauber E. Computer are social actors. 1994 Presented at: Conference on Hum Factors in Computing Systems; April 1994; Boston, MA p. 72-78. [doi: 10.1145/259963.260288]
76. Haug S, Castro RP, Kwon M, Filler A, Kowatsch T, Schaub MP. Smartphone use and smartphone addiction among young people in Switzerland. J Behav Addict 2015 Dec;4(4):299-307 [FREE Full text] [doi: 10.1556/2006.4.2015.037] [Medline: 26690625]

77. Kramer J, Künzler F, Mishra V, Smith SN, Kotz D, Scholz U, et al. Which Components of a Smartphone Walking App Help Users to Reach Personalized Step Goals? Results From an Optimization Trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
78. Haug S, Paz Castro R, Meyer C, Filler A, Kowatsch T, Schaub MP. A mobile phone-based life skills training program for substance use prevention among adolescents: pre-post study on the acceptance and potential effectiveness of the program, Ready4life. JMIR Mhealth Uhealth 2017 Oct 04;5(10):e143 [FREE Full text] [doi: 10.2196/mhealth.8474] [Medline: 28978498]
79. Haug S, Paz Castro R, Kowatsch T, Filler A, Dey M, Schaub MP. Efficacy of a web- and text messaging-based intervention to reduce problem drinking in adolescents: Results of a cluster-randomized controlled trial. J Consult Clin Psychol 2017 Feb;85(2):147-159. [doi: 10.1037/ccp0000138] [Medline: 27606700]
80. Schnyder MA, Santschi R. Auf Spurensuche im Ferienlager. Ein Lerncomic über Asthma. 2012. URL: https://www.lungenliga.ch/uploads/tx_pubshop/Lerncomic_Spurensuche.pdf [accessed 2020-04-13]
81. Zarouali B, Van den Broeck E, Walrave M, Poels K. Predicting consumer responses to a chatbot on Facebook. Cyberpsychol Behav Soc Netw 2018 Aug;21(8):491-497. [doi: 10.1089/cyber.2017.0518] [Medline: 30036074]
82. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digit Health 2019;5:2055207619880676 [FREE Full text] [doi: 10.1177/2055207619880676] [Medline: 31620306]
83. Filler A, Kowatsch T, Haug S, Wahle F, Staake T, Fleisch E. MobileCoach: A novel open source platform for the design of evidence-based, scalable and low-cost behavioral health interventions: Overview and preliminary evaluation in the public health context. : IEEE; 2015 Presented at: Wireless Telecommunications Symposium 2015 (WTS 2015); April 15-17, 2015; New York p. 1-6. [doi: 10.1109/wts.2015.7117255]
84. MobileCoach. URL: www.mobile-coach.eu [accessed 2021-02-06]
85. Stieger M, Wepfer S, Rüegger D, Kowatsch T, Roberts B, Allemand M. Becoming more conscientious or more open to experience? effects of a two–week smartphone–based intervention for personality change. Eur J Pers 2020 May 01;34(3):345-366. [doi: 10.1002/per.2267]
86. Haug S, Paz Castro R, Scholz U, Kowatsch T, Schaub MP, Radtke T. Assessment of the efficacy of a mobile phone-delivered just-in-time planning intervention to reduce alcohol use in adolescents: randomized controlled crossover trial. JMIR Mhealth Uhealth 2020 May 26;8(5):e16937 [FREE Full text] [doi: 10.2196/16937] [Medline: 32452818]
87. Lüscher J, Kowatsch T, Boateng G, Santhanam P, Bodenmann G, Scholz U. Social support and common dyadic coping in couples' dyadic management of type II diabetes: protocol for an ambulatory assessment application. JMIR Res Protoc 2019 Oct 04;8(10):e13685 [FREE Full text] [doi: 10.2196/13685] [Medline: 31588907]
88. Boateng G, Sels L, Kuppens P, Lüscher J, Scholz U, Kowatsch T. Emotion elicitation and capture among real couples in the lab. New York, NY: Association for Computing Machinery; 2020 Presented at: ACM CHI Conference on Human Factors in Computing Systems; 2020; Honolulu (cancelled).
89. Barata F, Tinschert P, Rassouli F, Steurer-Stey C, Fleisch E, Puhan MA, et al. Automatic recognition, segmentation, and sex assignment of nocturnal asthmatic coughs and cough epochs in smartphone audio recordings: observational field study. J Med Internet Res 2020 Jul 14;22(7):e18082 [FREE Full text] [doi: 10.2196/18082] [Medline: 32459641]
90. Nielsen J, Landauer TK. Model of the finding of usability problems. 1993 Presented at: ACM Conference on Human Factors in Computing Systems (CHI); 1993; Netherlands p. 206-213.
91. Max Dein Asthmacoach. URL: https://www.max-asthmacoach.ch/ [accessed 2021-02-06]
92. Falkenström F, Hatcher RL, Skjulsvik T, Larsson MH, Holmqvist R. Development and validation of a 6-item working alliance questionnaire for repeated administrations during psychotherapy. Psychol Assess 2015 Mar;27(1):169-183. [doi: 10.1037/pas0000038] [Medline: 25346997]
93. Kamis A, Koufaris M, Stern T. Using an attribute-based decision support system for user-customized products online: an experimental investigation. MIS Quart 2008;32(1):159-177. [doi: 10.2307/25148832]
94. Schnyder M. Begleitbericht Zur Entwicklung eines Lernmittels für Asthmakinder im Alter von 7 bis 12 Jahren. Universität Bern: Dissertation; 2014.
95. MAX - A Conversational Agent for Children with Asthma, Their Parents, and Healthcare Professionals - Data 2019. Open Science Framework. URL: https://osf.io/dpw9f/ [accessed 2021-02-06]
96. Schachner T, Keller R, V Wangenheim F. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. J Med Internet Res 2020 Sep 14;22(9):e20701 [FREE Full text] [doi: 10.2196/20701] [Medline: 32924957]
97. Ma T, Chattopadhyay D, Sharifi H. Virtual humans in health-related interventions: A meta-analysis. 2019 Presented at: Human Factors in Computing Systems; 2019; UK p. 1-6. [doi: 10.1145/3290607.3312853]
98. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]

99. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
100. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
101. Montenegro JLZ, da Costa CA, da Rosa Righi R. Survey of conversational agents in health. Exp Syst Apps 2019 Sep;129:56-67. [doi: 10.1016/j.eswa.2019.03.054]
102. van der Heijden H. User acceptance of hedonic information systems. MIS Quart 2004;28(4):695-704. [doi: 10.2307/25148660]
103. Rideout V. Measuring time spent with media: the Common Sense census of media use by US 8- to 18-year-olds. J Child Media 2016 Jan 18;10(1):138-144. [doi: 10.1080/17482798.2016.1129808]
104. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med 2019 May 16;9(3):440-447. [doi: 10.1093/tbm/ibz043] [Medline: 31094445]
105. Thompson D, Baranowski T. Chatbots as extenders of pediatric obesity intervention: an invited commentary on "Feasibility of Pediatric Obesity & Pre-Diabetes Treatment Support through Tess, the AI Behavioral Coaching Chatbot". Transl Behav Med 2019 May 16;9(3):448-450. [doi: 10.1093/tbm/ibz065] [Medline: 31094432]
106. Hirai K, Enseki M, Tabata H, Nukaga M, Matsuda S, Kato M, et al. Objective measurement of frequency and pattern of nocturnal cough in children with asthma exacerbation. Ann Allergy Asthma Immunol 2016 Aug;117(2):169-174. [doi: 10.1016/j.anai.2016.06.007] [Medline: 27372657]
107. Marsden PA, Satia I, Ibrahim B, Woodcock A, Yates L, Donnelly I, et al. Objective cough frequency, airway inflammation, and disease control in asthma. Chest 2016 Jun;149(6):1460-1466. [doi: 10.1016/j.chest.2016.02.676] [Medline: 26973014]
108. Barata F, Kipfer K, Weber M, Tinschert P, Fleisch E, Kowatsch T. Towards device-agnostic mobile cough detection with convolutional neural networks. 2019 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); 2019; China p. 1-11.
109. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: A pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219 [FREE Full text] [doi: 10.1037/hea0000306] [Medline: 26651462]
110. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
111. Künzler F, Mishra V, Kramer J, Kotz D, Fleisch E, Kowatsch T. Exploring the State-of-Receptivity for mHealth Interventions. Proc ACM Interactive, Mobile, Wearable and Ubiquitous Technologies 2019;3(4):1-27.
112. Heilbron FC, Escorcia V, Ghanem B, Niebles JC. ActivityNet: A large-scale video benchmark for human activity understanding. 2015 Presented at: 2015 IEEE Conference on Computer Vision and Pattern Recognition; June 7-12, 2015; Boston, MA p. 961-970.
113. Ramsey RR, Plevinsky JM, Kollin SR, Gibler RC, Guilbert TW, Hommel KA. Systematic review of digital interventions for pediatric asthma management. J Allergy Clin Immunol Pract 2020 Apr;8(4):1284-1293. [doi: 10.1016/j.jaip.2019.12.013] [Medline: 31870809]
114. Kramer J, Künzler F, Mishra V, Presset B, Kotz D, Smith S, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540 [FREE Full text] [doi: 10.2196/11540] [Medline: 30702430]

Abbreviations

BCT: behavioral change technique
CDHI: Centre for Digital Health Interventions
JITAI: just-in-time adaptive intervention

Edited by R Kukafka, G Eysenbach; submitted 16.10.20; peer-reviewed by E Bellei, A Bhatia; comments to author 09.11.20; revised version received 19.11.20; accepted 22.01.21; published 17.02.21.

Please cite as:
Kowatsch T, Schachner T, Harperink S, Barata F, Dittler U, Xiao G, Stanger C, v Wangenheim F, Fleisch E, Oswald H, Möller A Conversational Agents as Mediating Social Actors in Chronic Disease Management Involving Health Care Professionals, Patients, and Family Members: Multisite Single-Arm Feasibility Study
J Med Internet Res 2021;23(2):e25060
URL: http://www.jmir.org/2021/2/e25060/
doi:10.2196/25060
PMID:33484114

Original Paper

Hybrid Ubiquitous Coaching With a Novel Combination of Mobile and Holographic Conversational Agents Targeting Adherence to Home Exercises: Four Design and Evaluation Studies

Tobias Kowatsch[1,2,3], PhD; Kim-Morgaine Lohse[1], MSc; Valérie Erb[4,5], BSc; Leo Schittenhelm[2], MSc; Helen Galliker[1], BSc; Rea Lehner[3], PhD; Elaine M Huang[5], PhD

[1]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[2]Centre for Digital Health Interventions, Institute of Technology Management, University of St Gallen, St Gallen, Switzerland

[3]Future Health Technologies Programme, Campus for Research Excellence and Technological Enterprise, Singapore-ETH Centre, Singapore, Singapore

[4]Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

[5]People and Computing Lab, Department of Informatics, University of Zurich, Zurich, Switzerland

Corresponding Author:
Tobias Kowatsch, PhD
Centre for Digital Health Interventions
Department of Management, Technology, and Economics
ETH Zurich
WEV-G, Weinbergstrasse 56/58
Zurich, 8092
Switzerland
Phone: 41 712247244
Fax: 41 712247301
Email: tkowatsch@ethz.ch

Abstract

Background: Effective treatments for various conditions such as obesity, cardiac heart diseases, or low back pain require not only personal on-site coaching sessions by health care experts but also a significant amount of home exercises. However, nonadherence to home exercises is still a serious problem as it leads to increased costs due to prolonged treatments.

Objective: To improve adherence to home exercises, we propose, implement, and assess the novel coaching concept of hybrid ubiquitous coaching (HUC). In HUC, health care experts are complemented by a conversational agent (CA) that delivers psychoeducation and personalized motivational messages via a smartphone, as well as real-time exercise support, monitoring, and feedback in a hands-free augmented reality environment.

Methods: We applied HUC to the field of physiotherapy and conducted 4 design-and-evaluate loops with an interdisciplinary team to assess how HUC is perceived by patients and physiotherapists and whether HUC leads to treatment adherence. A first version of HUC was evaluated by 35 physiotherapy patients in a lab setting to identify patients' perceptions of HUC. In addition, 11 physiotherapists were interviewed about HUC and assessed whether the CA could help them build up a working alliance with their patients. A second version was then tested by 15 patients in a within-subject experiment to identify the ability of HUC to address adherence and to build a working alliance between the patient and the CA. Finally, a 4-week n-of-1 trial was conducted with 1 patient to show one experience with HUC in depth and thereby potentially reveal real-world benefits and challenges.

Results: Patients perceived HUC to be useful, easy to use, and enjoyable, preferred it to state-of-the-art approaches, and expressed their intentions to use it. Moreover, patients built a working alliance with the CA. Physiotherapists saw a relative advantage of HUC compared to current approaches but initially did not see the potential in terms of a working alliance, which changed after seeing the results of HUC in the field. Qualitative feedback from patients indicated that they enjoyed doing the exercise with an augmented reality–based CA and understood better how to do the exercise correctly with HUC. Moreover, physiotherapists highlighted that HUC would be helpful to use in the therapy process. The longitudinal field study resulted in an adherence rate of 92% (11/12 sessions; 330/360 repetitions; 33/36 sets) and a substantial increase in exercise accuracy during the 4 weeks.

Conclusions: The overall positive assessments from both patients and health care experts suggest that HUC is a promising tool to be applied in various disorders with a relevant set of home exercises. Future research, however, must implement a variety of exercises and test HUC with patients suffering from different disorders.

(J Med Internet Res 2021;23(2):e23612) doi:10.2196/23612

KEYWORDS

ubiquitous coaching; augmented reality; health care; treatment adherence; design science research; physiotherapy; chronic back pain; pain; chronic pain; exercise; adherence; treatment; conversational agent; smartphone; mobile phone

Introduction

Musculoskeletal disorders (MSDs), such as rheumatoid arthritis, osteoarthritis, low back pain, neck pain, or gout, negatively impact the locomotor system and are influenced by various risk factors, such as a sedentary lifestyle, malnutrition, or obesity [1]. Individuals with MSDs suffer from chronic pain, impaired mobility and physical function, and reduced quality of life [2]. The average estimated global prevalence of MSDs lies at 18% [3] and increases with age [4]. MSDs account for 21.3% of all years lived with disability worldwide [5], with back pain as the leading cause of disability [3]. Therefore, both the individual's psychosocial burden and the financial burden of MSDs are significant. In the United States, for example, treatment costs are estimated to account for 5.7% of gross domestic product [6]. As a consequence of population growth, aging, and sedentary lifestyles, the burden of MSDs will dramatically increase in the future and thus poses significant challenges to global health [4,5]. It is, therefore, crucial to develop effective interventions for individuals with MSDs.

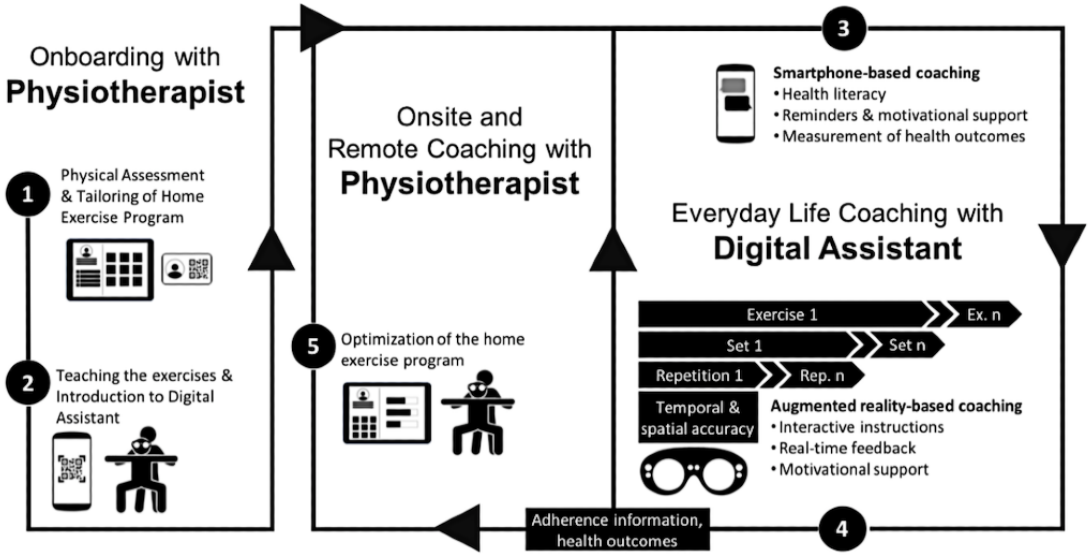
Physiotherapy is an effective intervention for MSDs [7-10]. However, treatment adherence (ie, the extent to which an individual's behavior "corresponds with agreed recommendations from a health care provider" [11]) is a common problem in various health care settings [11-13], including home exercises in physiotherapy [11,14-17]. For example, treatment adherence ranges from 60% [18] down to 30% [9,19,20], resulting in increased costs because of prolonged treatment [5,8,21-23]. Common treatment adherence dimensions (TADs) [24-26] are the session completion rate (the number of completed vs prescribed exercise sessions, TAD1), the set completion rate (the number of completed vs prescribed sets for each exercise, TAD2), the exercise repetition rate (the number of completed vs prescribed exercise repetitions within each set, TAD3), the temporal exercise accuracy (the actual vs prescribed velocity of exercise execution, TAD4), and the spatial exercise accuracy (the actual vs prescribed body movement trajectory, TAD5). Reasons for nonadherence can vary. For example, patients may simply forget to follow the various TADs, and further patient-related factors, such as disability, attitudes, motivation, and beliefs about exercise risks and benefits, can also negatively impact the TADs [14,27-33].

Moreover, adherence support is often limited to on-site physiotherapy sessions, leaving patients alone at home with standardized exercise instructions, which results in low adherence rates [34-36]. Therefore, various technical approaches have been proposed to help increase home exercise adherence in physiotherapy [37]. A smartphone app, for example, motivates patients and sends reminder messages to individuals with an MSD [38], or a virtual reality training is used to provide feedback on exercise execution [39]. However, two recent reviews indicate that adherence remains a challenge [40,41]. Our review of remote patient monitoring tools and current scientific work (Multimedia Appendix 1 [16,29,39,42-66]) suggests that there is untapped potential in addressing this problem, too.

To this end, we propose hybrid ubiquitous coaching (HUC) to improve home exercise adherence for treatments that require not only personal on-site coaching sessions by health care experts but also a significant amount of home exercises. HUC relies primarily on a conversational agent (CA) that delivers relevant health literacy information about the importance and benefits of home exercises and personalized motivational exercise reminders via a smartphone. Moreover, the CA of HUC delivers real-time exercise support, monitoring, and feedback in a hands-free augmented reality (AR) environment. In HUC, health care experts introduce the CA as their trusted personal assistant that lives not only in the patient's pocket/smartphone in the form of a text-based physiotherapy chatbot [48,49,67], but also in the patient's AR glasses in the form of an embodied, holographic instructor [52,55] (see Figure 1 for an overview of HUC).

HUC combines for the very first time (to the authors' best knowledge) research on CAs, human-supported digital interventions, health psychology, and mobile and wearable technology–supported physical exercises. First, HUC employs a CA because they are able to build a working alliance with patients [49,57,68], which is an important relationship quality [69] that is robustly linked to treatment success [59]. Accordingly, there is a large body of evidence on the effectiveness of CAs in delivering clinical and nonclinical interventions [52]. Second, HUC relies on a hybrid coaching concept consisting of CA and human physiotherapist teams because digital interventions supported by humans result not only in higher treatment adherence [64,66], but also in better treatment outcomes [70,71]. Third, HUC offers real-time feedback about intervention progress and correct exercise execution, which, in turn, is assumed to increase self-efficacy [29], an important construct that helps shape positive attitudes toward therapy adherence and health-promoting behavior [63]. Finally, HUC aims to increase adherence by seamlessly delivering an intervention into the everyday lives of vulnerable individuals or patients with the help of mobile and wearable technology (ie, smartphones and AR glasses) [72,73].

Figure 1. Overview of hybrid ubiquitous coaching (HUC).



Because this research is unprecedented, the following research questions (RQs) were formulated to guide the design, implementation, and evaluation of HUC: (RQ1) How is HUC perceived by (a) patients and (b) health care experts?; and (RQ2) Does HUC lead to treatment adherence to home exercises?

To answer these questions, we applied HUC exemplarily to the field of physiotherapy. During the last 3 years, 4 design-and-evaluate loops were conducted: two studies were carried out in the lab with 50 physiotherapy patients, 11 physiotherapists were interviewed, and finally, empirical data from 1 patient during a 4-week longitudinal field study were collected and assessed to reveal real-world benefits and challenges related to HUC.

Methods

Overview

To address the challenge of nonadherence to home exercises in physiotherapy, we started the interdisciplinary development of HUC in collaboration with 2 physiotherapists. In a first step, a storybook was developed (Multimedia Appendix 2). In a second step, HUC was developed further as outlined in the next section. After the conceptual work, we formulated the research questions that guided the 4 studies in which we iteratively improved, implemented, and evaluated HUC. Table 1 provides an overview of the 2-year design and evaluation process.

Table 1. Overview of the studies.

Characteristic	Study 1, 2018	Study 2, 2018/2019	Study 3, 2019	Study 4, 2019
Maturity of HUC[a]	Version 1: AR[b]-based CA[c] coaching. Superhero exercise instructor and humanlike female exercise model.	Version 2: Revised AR-based CA coaching. One humanlike exercise instructor with professional physiotherapist clothing and additional smartphone-based CA coaching (mockup).		Version 3: Revised AR-based CA with dynamic behavior over time and Wizard of Oz smartphone-based CA coaching
Study design	Lab study	Semistructured interview and survey	Within-subject lab study	Observational longitudinal field study
Participants	35 patients	11 physiotherapists	15 patients, 2 physiotherapists	1 patient-physiotherapist dyad, 3 physiotherapists
RQ1a: How is HUC perceived by patients?	Perceived usefulness, ease of use & enjoyment.[d] Suggestions for improvements.[e]	N/A[f]	Patient-CA working alliance.[d] Perceived usefulness, ease of use, enjoyment, task load, exercise difficulty.[d,e]	Perceived usefulness, ease of use & enjoyment. Intention to continue using HUC, suggestions for improvements.[d]
RQ1b: How is HUC perceived by health care experts?	N/A	Patient-physiotherapist working alliance, relative advantage of HUC.[d,e]	N/A	Perceived usefulness and relative advantage of HUC, suggestions for improvements.[e]
RQ2: Does HUC lead to treatment adherence to home exercises?	Behavioral intention & recommendation to use HUC.[e]	Session completion rate, sets & repetition (TAD1-3[g])[d]	Spatial & temporal accuracy (TAD4-5)[d]	Session and set completion rates, exercise repetition rate, spatial & temporal accuracy (TAD1-5)[d]

[a]HUC: hybrid ubiquitous coaching.

[b]AR: augmented reality.

[c]CA: conversational agent.

[d]Quantitative feedback.

[e]Qualitative feedback.

[f]N/A: not applicable.

[g]TAD: treatment adherence dimension.

Concept of Hybrid Ubiquitous Coaching

A conceptual overview of HUC is illustrated in Figure 1 and outlined in the following paragraphs.

Onboarding With a Physiotherapist

First, a physiotherapist performs a physical assessment of the patient and, depending on the results, defines a tailored home exercise program. Next, the physiotherapist works together with the patient on the exercises from the program by focusing on the TADs (Table 1). Then, the physiotherapist introduces the CA as his or her scalable personal assistant that will support the patient regarding the TADs in their everyday life via the smartphone and the AR glasses. For this purpose, the physiotherapist shows the patient how to interact with the CA on the smartphone and via the AR glasses. To link the CA with the patient (eg, name, age, height, weight, diagnosis) and his or her specific home exercise program, the physiotherapist prints out a therapeutic prescription card with a quick response (QR) code. The QR code not only contains encrypted links to the patient data and personalized exercise program, but also data about the physiotherapist so that the CA "knows" the name of the physiotherapist it was linked to. This allows the CA to link back to the "real" world during conversations with the patient and strengthen the relationship and the working alliance with the specific physiotherapist. A strong working alliance between health professionals and patients establishes an attachment bond [74], the development of a shared understanding about goals and tasks, and, in return, boosts treatment adherence [69,75]. On the way home, the patient scans the QR code with his or her smartphone to download the mobile HUC app and start the interaction with the CA. The CA introduces itself as the physiotherapist's personal assistant and gives an overview of the upcoming digital coaching program. During this dialogue with the CA, the shipment of the AR hardware to the patient's home is triggered as the patient's address is made available via the link in the QR code.

Everyday Coaching

The smartphone- and AR-based CAs allow for remote everyday coaching. The combination of the two CAs together results in a richer communication channel that enables more accurate and personalized information to be transmitted to the patient [76,77]. The design of the HUC elements was also informed by behavioral change techniques (BCTs, ie, evidence-based intervention components aiming to change behavior) [78]. The smartphone-based CA reminds the patient to do the home exercises (TAD1), motivates the patient via the smartphone, and emphasizes the benefits of performing the exercise by providing psychoeducational material, for example, about sleep quality and degree of pain (BCTs 2, 13, 15, and 20 [78]). The AR-based CA delivers real-time exercise support by

demonstrating the execution (BCT 22), whereby arrows highlight important angles to be aware of and motivate the user during the exercise (BCT 13). The AR-based CA guides the patient through the exercise, monitoring the progress (TAD1-5, BCT 17) and giving real-time feedback (TAD4-5, BCT 19). It also informs and emphasizes essential aspects of how to perform the exercise correctly (TAD4-5, BCT 21). The AR-based CA counts the number of sets and repetitions out loud and provides feedback on the exercise execution after a completed set of exercises (TAD2-5, BCT 19). The feedback is both visual and auditory and is based on comparing patient data from the AR system with data about the physiotherapist performing the same exercise. Lastly, the AR-based CA aims at increasing the attachment bond, an important relationship quality and dimension of a working alliance [79], by literally giving high fives or showing the patient a funny dance move (BCT 13) when a set or training session is completed.

On-site and Remote Coaching With a Physiotherapist

HUC grants physiotherapists access to patients' data about the TADs via a web-based dashboard. Based on that data, physiotherapists can tailor and optimize planned on-site and remote coaching sessions more efficiently. For example, if squats are performed incorrectly and the real-time feedback from the AR-based CA is not processed correctly by patients, pressure could be applied on the knees rather than on the muscles, which could cause knee injuries. If this happens, the physiotherapist would get an alert message from the CA to target the weakness directly in the next online or on-site coaching session.

In the remainder of this paper, we describe 4 build-and-evaluate loops that we conducted to build and assess HUC according to our research questions. Table 1 provides an overview of this process.

Study 1: Perceptions of Individuals With Physiotherapy Experience

To address RQ1a and RQ2, we implemented a first version of HUC with a focus on the AR-based CA and tested it with 35 individuals with physiotherapy experience at a public fair in September 2018.

Development of HUC Version 1

HUC included an AR-based CA that appeared as a male cartoonlike superhero assistant of a human male physiotherapist and a female human-looking exercise model, demonstrating the exercise execution (Figure 2). A personal, empathetic, and humanlike CA can increase a patient's intention to change behavior and to continue to use the CA [80]. Further, the rationale behind the superhero design was to make the user experience more fun. The rationale for the design features was to design a personal and empathetic CA. For ease of presentation, a squat exercise was implemented as it is a common physiotherapy exercise. After the CA demonstrated the exercise (eg, by flying around the female character and pointing out important aspects of the squat movements) and commented via voice on how to correctly execute the squat, the CA asked the patient to perform the exercise. Patients were able to respond to voice-based questions from the CA with predefined answer options that appeared as speech bubbles in the AR space around 20 cm in front of and at eye level for the patient (Figure 3). The patient was able to select an answer by touching it with a finger. This hands-free approach does not require an additional controller and thus allows for more intuitive and direct interactions [81] (Figure 4). A significant advantage of this is the direct manipulations and interactions in the AR space. Interactions were primarily used to increase the attachment bond via small talk (eg, "How are you today?"), to explain the exercise (eg, "Let's have a close look at Alexis' movements"), or to progress through the several-step process of exercise execution ("well done, four more to go!"). See Multimedia Appendix 3 for the video clip and Multimedia Appendix 4 for the technical details of the implementation and hardware.

Figure 2. Augmented reality–based conversational agent and female humanlike model demonstrating the squat exercise.

Figure 3. Predefined answer options. The conversational agent communicated visually and auditorily.

Figure 4. Participant performing the squat exercise while wearing the augmented reality hardware.

Evaluation

Participant Acquisition

Participants were recruited from a public fair using convenience sampling [82]. They were included if they had already participated in physiotherapy sessions and were interested in HUC.

Procedure

Participants who expressed an interest in HUC and provided their consent to participate in the study were invited to participate. HUC was then introduced, and usage of the AR glasses was demonstrated. Next, participants were asked to start the interaction with the AR-based CA and performed one exercise session. Afterward, participants were asked to fill out a feasibility questionnaire. Participants received no monetary compensation.

Measurements

HUC was assessed based on technology acceptance [83-85] and word-of-mouth [86] research. To reduce participant burden, and due to the feasibility character of this first study, we used single-item measures for perceived enjoyment [87], perceived ease of use, perceived usefulness [85], and intention to use [83]. Consistent with prior work [87], 7-point Likert scales were used, ranging from strongly disagree (1) to strongly agree (7). Further, the Net Promoter Score (NPS) [86] was used to assess whether participants would recommend HUC to other patients. The NPS is a single-item measurement that indicates satisfaction with a service. NPS scores are binned into four categories: −100 to 0

(needs improvement), 1 to 30 (good), 31 to 70 (great), and 71 to 100 (excellent). Finally, qualitative feedback was gathered on positive aspects of HUC, and suggestions for improvement were provided by participants.

Study 2: Revision of HUC and Evaluation by Physiotherapists

To address RQ1b and RQ2, we revised HUC based on the feedback from study 1 and assessed it with 11 physiotherapists between November 2018 and March 2019.

Design and Technical Implementation

HUC was revised as follows based on the qualitative results of study 1. The AR-based CA was scaled up to be human-sized and adapted in appearance to look like a physiotherapist (ie, the clothes, including the logo, were copied from a real physiotherapist, since a professional look and a natural humanlike style are agent characteristics associated with increased intention to use [80]). Moreover, due to the limited field of view of the AR glasses, which sometimes made it hard to see the interactions between the two characters in HUC version 1 and thus confused some of the participants in study 1, the female humanlike model was omitted. Therefore, in study 2 the revised AR-based CA demonstrated the squat exercise with the help of his own virtual body (Figure 5 and the video in Multimedia Appendix 5). Moreover, an exercise tutorial was added to help participants better understand how to correctly execute the exercise (BCT 22). For this purpose, virtual guides indicated important aspects of exercise execution (Figure 5). Furthermore, automatic error detection was implemented in collaboration with 2 physiotherapists. For the error detection, the position and rotation of the headset were tracked and compared with the squat movements of the physiotherapists (Multimedia Appendix 4). Participants were asked to position themselves on virtual footprints placed on the floor and to remain in that position during the course of the exercise. Thus, participants' initial position and rotation were determined. Additionally, participants' heights were stored as the difference between the initial position and the floor along the vertical axis. Based on these parameters, errors for insufficient depth of the squat, too fast or too slow execution, and too much deviation to the left, right, front, or back from an individual's vertical center axis were detected. Further technical details of the error detection are outlined in Multimedia Appendix 4. The errors detected were saved in the AR app so that the CA would be able to generate and send an error report to the corresponding physiotherapist. For example, if a participant did not move low enough with the upper body during a specific squat exercise, the AR-based CA would say, "A little bit lower." In addition to this error-related real-time feedback, and to target TAD1-3, the AR-based CA took over the moderation during the exercises with motivating voice-based instructions and progress reports (BCTs 12, 13, and 15). For example, the AR-based CA counted up the number of repetitions (eg, "Only three… two… one. Great! You already finished the 1st set, take a quick break and then let's start with the 2nd set of squats"). Moreover, real-time feedback was implemented based on these errors and communicated during and after a specific exercise, thus targeting temporal (TAD4) and spatial (TAD5) accuracies.

Finally, the HUC concept outlined in Figure 1 was broken down into a flowchart diagram and specific sketches (eg, of the web-based dashboard for physiotherapists and the smartphone-based CA interaction; Multimedia Appendix 6) that illustrated the various intervention components of HUC in more detail. In addition to the revision of the AR-based exercise, the flowchart and sketches were then used for the assessment of the various aspects of HUC by physiotherapists as outlined below.

Figure 5. Virtual guides embedded into the revised augmented reality–based conversational agent and used in studies 2, 3, and 4.

Evaluation

Participant Acquisition

Physiotherapists treating patients with MSDs were recruited using chain sampling [82] until saturation of the qualitative feedback was reached [88].

Procedure

First, we explained the HUC concept, the flowchart, and the sketches to each participating physiotherapist. Then, the physiotherapist performed the specific squat exercise session with the AR-based CA. Thereafter, a semistructured interview was conducted to gather feedback on the utility and feasibility of HUC. The interview was recorded and transcribed according to the rules of simple transcription [89]. Thereafter, relevant themes and concepts were extracted following an iterative coding process [90]. After the interview, we sent an online survey to the physiotherapist in which we asked them to assess the relative advantage of HUC compared to current patient monitoring and communication applications and asked about the potential of HUC to strengthen the working alliance between the physiotherapist and their patient. No monetary compensation was provided.

Measurements

The guiding questions of the semistructured interview are listed in Multimedia Appendix 7. The perceived relative advantage of HUC was adapted from prior work [91]. The answer options on the 6-item instrument were anchored on 7-point Likert scales, ranging from strongly disagree (1) to strongly agree (7). The Session Alliance Inventory [92] was used to assess the working alliance. Answer options on the 6-item instrument were anchored on 7-point Likert scales, ranging from never (1) to always (7).

Study 3: Revised HUC (Version 2) Assessed by Patients

To address RQ1a and RQ2, HUC was assessed by 15 patients seeking physiotherapy treatment in January and February 2019. To assess the relative advantage of HUC to commonly employed methods of exercise instruction, HUC was compared to paper-based and video-based exercise instruction.

Design and Technical Implementation

The revised HUC (version 2) as outlined in study 2 above was used for the assessment.

Evaluation

Participant Acquisition

Patients were recruited from a physiotherapy center of one of the collaborating physiotherapists. Inclusion criteria were age ≥18, participation in at least 3 physiotherapy sessions, no experience with the squat exercise, being in the physical condition to perform squat exercises, normal vision or wearing contact lenses, and normal hearing.

Procedure

A two-period crossover study design was used to assess the relative advantage of HUC compared to paper-based and video-based exercise instruction, two commonly used methods in physiotherapy. This study design also allowed us to account for learning effects [93]. The study was conducted at the physiotherapist center. After giving consent, each participant was systematically assigned to a specific order of the exercise instruction method. Then, a physiotherapist described and demonstrated the squat exercise in the therapy session and ensured correct exercise execution. Next, patients received either handwritten paper-based exercise instructions (Figure 6), video-based instructions (Figure 7 and Multimedia Appendix 8), or a QR code on a business card to download the HUC app and the AR hardware (Figure 5 and Multimedia Appendix 5). The content of the exercise instruction was consistent for all three instruction methods. Each patient was then instructed to perform 3 sets of squats with 10 repetitions each. The exercises were video-recorded so that 2 physiotherapists were able to assess the accuracy of the exercise execution independently from each other at a later point in time (see Measurements). After each of the 3 exercise sessions, patients were asked to assess the instruction method via an online survey. Finally (ie, after all exercise sessions), a semistructured interview was conducted with each participant to gather suggestions to improve HUC. The overall duration of approximately 1 hour was compensated with US $50.

Figure 6. Paper-based instructions of study 3.

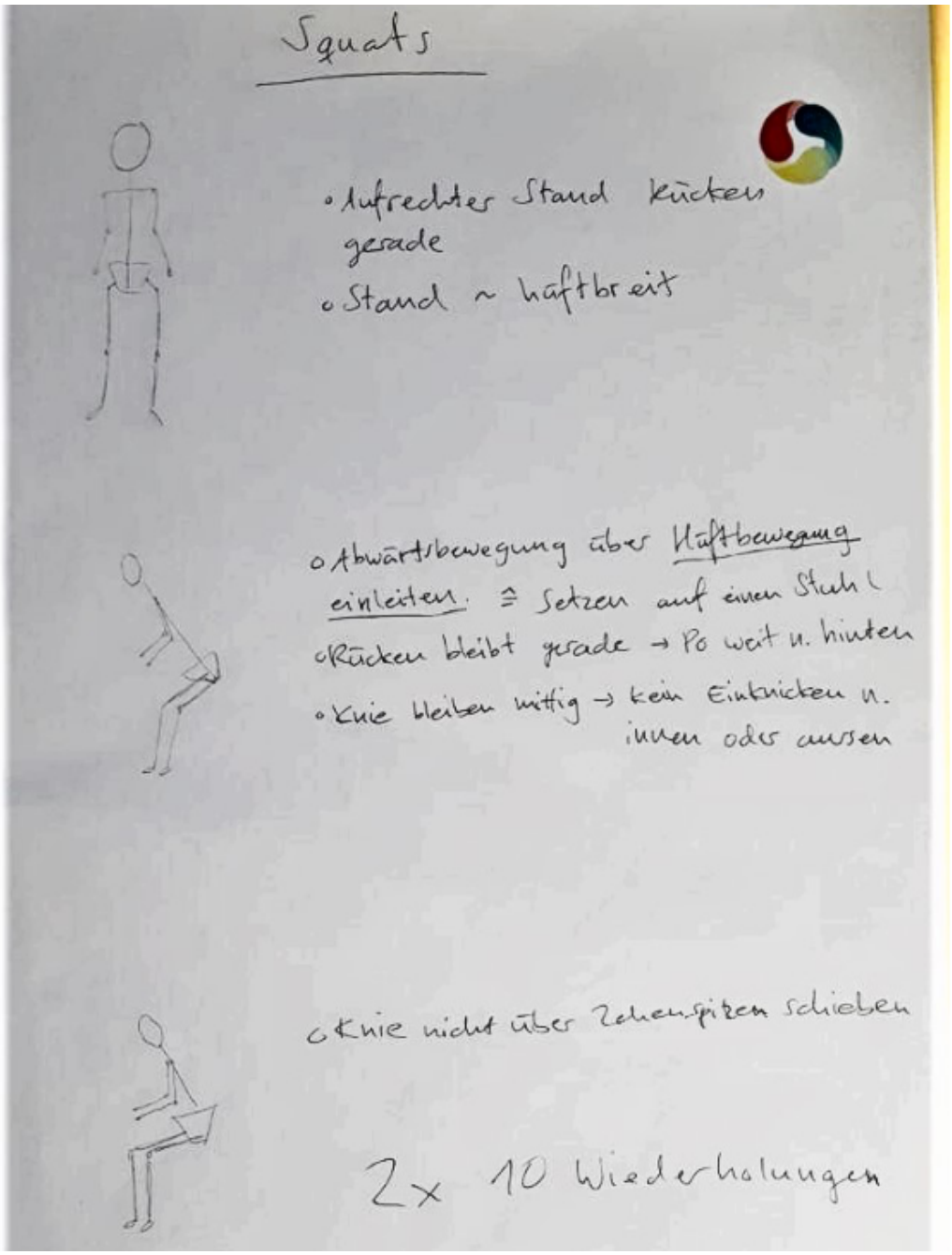


Figure 7. Video-based instructions of study 3.

Measurements

The exercise instructions were assessed with mental effort and frustration items from the task load index [94]. Moreover, perceived enjoyment, ease of use, and usefulness of the exercise instructions were adopted from technology acceptance research [85,95]. We also assessed the working alliance between the patient and HUC's AR-based CA with the Session Alliance Inventory [92]. Additionally, the intention to continue working with the corresponding exercise instruction was determined based on prior work [96]. We also asked patients to rank the three instruction exercises according to their informativeness. We finally adopted 5 assessment items from the Exercise Assessment Scale [97]. With these items, 2 independent physiotherapists were asked to assess the correctness of the exercise execution for each instruction method and participant with a maximum score of 6 points. For this purpose, patients performing the squat were video-recorded from two perspectives (front and side). All quantitative survey items are listed in Multimedia Appendix 9. The questions of the semistructured interview focused on patients' perceptions and suggestions for improvements (Multimedia Appendix 10).

Study 4: HUC Assessed in a 4-Week Field Study

Due to the high burden nature of the study, with experimental technology that required support from the researchers, we conducted a 4-week n-of-1 trial to complement our findings from studies 1-3 and to answer RQ1a, RQ1b, and RQ2. The goal of this study was to complement the findings from the previous studies and to add external validity to the lab results by assessing the long-term adherence and feasibility of HUC in the everyday life of one patient. By assessing HUC with only one patient, this study was not designed to illustrate the breadth of possible use experiences but to show one experience in depth and thereby potentially reveal real-world benefits and challenges [98,99].

Design and Technical Implementation

The objective of the third design loop was to adapt HUC for long-term use. For the field study assessment of HUC, there was no QR code provided, and no dedicated HUC was developed. Consistent with recent work [80,100,101], the following revisions were implemented to increase variation in HUC coaching.

First, depending on the patient's familiarity with the AR-based app (eg, how often it was used), tutorials became skippable, and explanations were shortened after the first week. Second, there was randomization of the coach's speaking texts ("Hey, good to see you again!"), animations (eg, stretching the upper limb), and interactions (eg, clapping, high five, thumbs up). Third, randomization of the coach's texture (ie, the AR-based CA had a set of different colored clothes) and the background music (eg, energetic rock and funk) changed. Lastly, the AR-based app showed the patient the number of sessions already completed as an early form of progress visualization, a well-established behavioral change technique (BCT 19) [78].

Evaluation

Participant Acquisition

A new physiotherapy patient who did not participate in one of the other studies was recruited from a physiotherapy center. The eligibility criteria were identical to those outlined in study 3. Additionally, 3 physiotherapists from the same center were recruited for a follow-up interview. One physiotherapist was the therapist treating the patient, the second therapist was involved in the evaluation in the course of the third study, and the third therapist was not involved in any of the previous studies.

Procedure

Before starting the study, the patient signed the informed consent. Then, at the beginning of the study, the patient received a study information sheet from the physiotherapist and agreed to practice a squat exercise 3 times per week. HUC was then demonstrated, and the patient received an introduction to the AR headset. Thereafter, the patient was coached by HUC in their everyday life as outlined in Figure 1. Again, the exercise in focus was the squat. The coaching included, on average, 4 weekly text messages via WhatsApp, aimed to elicit reminders, to provide psychoeducational material, and to motivate the patient to perform the squat as often as the physiotherapists recommended (BCTs 2, 13, 15, and 20 [78]). According to the Wizard-of-Oz method [102], the CA was simulated by a coauthor without the patient being aware of it and reminded the patient weekly to conduct the exercises over a 4-week period. Further, real-time feedback was provided during the execution of the exercises (Multimedia Appendix 5). The TAD data (eg, session completion rate, sets and repetitions, errors) were collected and sent to a coauthor via email so that additional feedback could be provided via a WhatsApp-based text message (eg, to motivate the patient to perform an exercise). After the trial, the patient was invited to take part in a debriefing interview, during which the experience with HUC was explored and the patient was asked for suggestions for improvement. The patient received a monetary compensation of US $50 for participating. In the case of any therapeutic or technical questions, the patient was able to contact the physiotherapist and the study team. Finally, the results of the n-of-1 trial were shown to 3 physiotherapists to gather additional qualitative feedback about HUC (eg, suggestions for improvement).

Results

The results of the 4 studies are presented in the following sections. The depersonalized raw data and data analysis script are made available in Multimedia Appendices 9 and 11, respectively.

Study 1: Perceptions of Individuals With Physiotherapy Experience

Overall, 35 (13 females) individuals with a mean age of 35 years (SD 11) and previous physiotherapy experience participated in the study. Most of them (32, 91%) had experience with the exercise. The results listed in Table 2 indicate that participants enjoyed the exercise sessions and perceived the exercise moderated by the AR-based CA to be useful and easy to use. Moreover, participants indicated that they would be willing to use this form of AR-based exercise at home. All of these assessments lie significantly above the neutral scale value by conducting Wilcoxon signed rank tests (Table 2). The NPS was negative and close to zero, indicating that the HUC needs improvement before participants would recommend it to other patients. This result was expected given the prototype character of this first version of HUC.

Table 2. Augmented reality–based conversational agent coaching assessed by 35 patients.

Construct	Items	Mean[a] (SD)	*P* value[b]
Perceived enjoyment	I enjoyed the exercise with Alex.[c]	5.74 (1.06)	<.001
Perceived ease of use	It was easy to follow the exercise instructions.	6.14 (0.84)	<.001
Perceived usefulness	I was able to follow the exercise.	6.37 (0.80)	<.001
Intention to use	I would use this type of holographic exercise at home.	5.40 (1.49)	<.001
Net Promoter Score[d]	How likely is it that you would recommend this type of exercise to other patients?	−17.14[e]	N/A[f]

[a]7-point Likert scales ranging from strongly disagree (1) to strongly agree (7).

[b]Wilcoxon signed rank test with test value 4.

[c]The augmented reality–based conversational agent was given the name Alex.

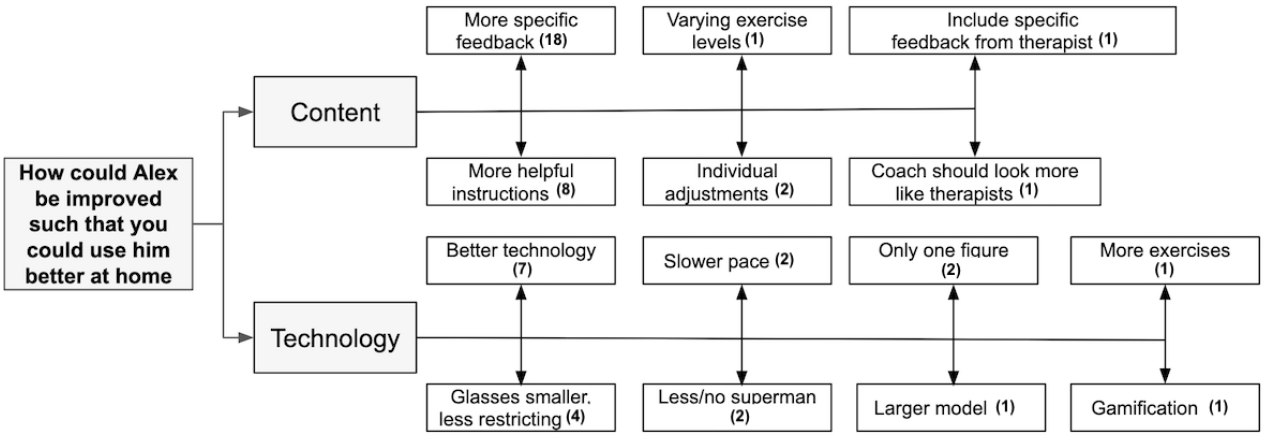
[d]NPS ranges between −100 and 100.

[e]The percentage of detractors subtracted from the percentage of promoters.

[f]N/A: not applicable.

The qualitative feedback on the positive aspects of HUC and suggestions for improvement are shown in the thematic maps (Figure 8). Suggested improvements included “more specific feedback,” “more helpful instructions,” “coach should look more like a physiotherapist,” “only one figure,” and “less/no superman.”

Figure 8. Thematic map of the patients' qualitative feedback of study 1. Note: the number in brackets indicates the frequency a topic was mentioned.



Study 2: Evaluation by Physiotherapists

Eleven physiotherapists with 3 to 30 years of experience (6 females, age of both sexes between 20 and 49 years) participated in the study. Five physiotherapists worked in a private practice, 4 in a hospital, and 2 in a rehabilitation clinic.

The physiotherapists indicated in the interviews that patients could be better supported with the HUC and that physiotherapists would intend to use it for treatment sessions. The AR-based CA was perceived as an additional motivator due to its personal, interactive, and playful approach. The exercise guidance by the life-sized CA, including the real-time feedback (eg, automatic counting of exercise repetitions), was also perceived as an improvement to the status quo. The smartphone-based CA was perceived as an advantage over current solutions, mainly due to its ability to remind patients to execute the exercise, but also by providing psychoeducational content that targets increasing patients' health literacy. Physiotherapists were, however, skeptical of whether HUC could improve the quality of the treatment and adherence. The HUC was seen to have some potential to strengthen the working alliance between physiotherapists and their patients. Establishing therapeutic goals was seen as an advantage, since tracking enables patients to be continuously informed about the degree of goal achievement. The ability to instruct and provide feedback on the exercise technique was also identified as a factor that could potentially improve the shared understanding of treatment tasks. However, the majority of the physiotherapists expressed the belief that mutual trust, empathy, and, consequently, mutual goals and tasks were uniquely established during face-to-face encounters and that technological systems could not successfully act as assistants to foster a good working alliance. Lastly, physiotherapists suggested that continuously monitoring a patient could have a negative influence on mutual trust.

This qualitative feedback is supported by the quantitative data from the online survey (Table 3). The physiotherapists saw a clear relative advantage of HUC relative to the status quo of commercial applications. The working alliance assessments also resulted in a confirmation of the qualitative feedback. Overall, the experts assessed the support of the CA as not being sufficient to build and maintain a robust working alliance between patients and physiotherapists. This was especially the case for establishing an attachment bond.

Table 3. HUC assessed relative to commercial applications by 11 physiotherapists.

Construct	Exemplary item	Items, n	Mean[a] (SD)	*P* value
Relative advantage	With HUC[b], patients would perform the exercises more correctly compared to previous methods.	6	4.97 (0.96)	<.001[c]
Working alliance				
Goal agreement	Alex would help me (my patient and I) work towards mutually agreed goals.	2	4.23 (1.31)	.19[d]
Task agreement	Alex would help me convince the patient that the way we work on the problem is correct.	1	3.82 (1.47)	.64[d]
Attachment bond	Alex would help me make the patient appreciate me.	3	2.06 (1.32)	.99[d]
Total		6	3.08 (1.68)	.99[d]

[a]7-point Likert scales ranging from strongly disagree (1) to strongly agree (7) were used.

[b]HUC: hybrid ubiquitous coaching.

[c]*t* test (2-tailed; t_{10}=8.204) with test value 4.

[d]Wilcoxon signed rank test with test value 4.

Study 3: Revised HUC (Version 2) Assessed by Patients

Fifteen patients (9 female) with a mean age of 37 years (SD 9.93) participated in the study. Participants were generally very positive toward HUC, the video instruction, and, to some extent, the paper instructions (Table 4). In total, 8 patients (53%) rated HUC as their favorite training method, while 6 patients (40%) rated the video instructions as their favorite method. Further, in the HUC method, the overall working alliance, task agreement, and attachment bond values were statistically significantly higher than the neutral scale value of 4. HUC resulted in higher average scores on the adapted exercise assessment scale (TAD4-5) compared to the video- and paper-based instructions.

The qualitative feedback suggested that the three instruction methods provided different levels of richness of information. Eight patients described the paper exercises as boring and not enjoyable to follow. Even though patients found the paper instructions helpful to recall the exercise, they often had trouble interpreting the instructions correctly. Three patients found the video instructions authentic and found the exercise easy to understand. However, patients mentioned that they felt insecure about the correctness of exercise execution. Patients also found the exercises easier to understand while using HUC and appreciated the guidance, the reminder app, and the real-time feedback to understand when mistakes were made. Eight patients also perceived HUC as fun and enjoyable. Six patients particularly enjoyed performing the exercise together with a life-sized CA. One patient, however, found that the CA did not engage emotionally. A disadvantage of HUC perceived by 9 patients—thereby a frequently perceived disadvantage—was the bulkiness of the AR hardware used in this study.

The patients suggested several improvements (Figure 9). First, 3 patients wanted to personalize the CA (eg, adjust the speed and the movement parameters to account for varying abilities, individually decide the appearance of the coach). Second, 2 patients mentioned a preference for a real person and not an animated character (eg, *"looking a bit more human-like, not such a computer-figure").* Third, 2 patients suggested adding more detailed real-time feedback and reminders and elements of gamification (eg, rewards for regularity, real-time reminders, and accuracy). Overall, HUC was positively perceived.

Table 4. HUC-based, video-based, and paper-based exercise instruction methods assessed by 15 physiotherapy patients and exercise assessment scale assessed by 2 physiotherapists. Note: _ represents the exercise instruction method.

Construct	(Exemplary) item	n	HUC[a] M[b] (SD)	*P* value[c]	Video M (SD)	*P* value[c]	Paper M (SD)	*P* value[c]
Task load linked to exercise instruction method[d]								
Mental capacity	How much mental and perceptual activity was required with _?	1	2.66 (1.11)	.005	3.06 (1.43)	.04	4.00 (1.36)	.75
Frustration	How frustrated did you feel during the execution with the _?	1	2.66 (1.44)	.005	2.60 (1.24)	.004	2.80 (1.69)	.02
Perceived characteristics of the instruction method[e]								
Perceived enjoyment	_ was fun to use.	1	5.26 (1.90)	.06	5.26 (1.09)	.005	4.13 (1.06)	.91
Perceived ease of use	I could understand _ very easily.	1	6.20 (1.01)	.001	5.80 (1.08)	.002	5.53 (1.40)	.005
Perceived usefulness	_ helped me to do the exercises correctly.	3	5.59 (1.63)	<.001	5.07 (1.52)	<.001	4.76 (1.42)	.002
Patient–conversational agent working alliance[e]								
Goal agreement	Alex and I agree on what is important for me to work on.	2	4.89 (1.96)	.06	N/A[f]	N/A	N/A	N/A
Task agreement	The way Alex and I are working with my problem is correct.	1	5.14 (1.51)	.02	N/A	N/A	N/A	N/A
Attachment bond	Alex and I respect each other.	3	5.57 (1.45)	<.001	N/A	N/A	N/A	N/A
Total		6	5.27 (1.65)	<.001	N/A	N/A	N/A	N/A
Intention to continuously use[e]	How much would you like to continuously use the [instruction method]?	1	5.07 (1.94)	.10	5.14 (1.29)	.01	5.14 (0.86)	.002
Exercise assessment scale[g]	Correct body part moving in correct plane	4	4.23 (0.75)	<.001	4.17 (1.06)	<.001	3.63 (0.90)	<.001

[a]HUC: hybrid ubiquitous coaching.

[b]M: mean.

[c]Wilcoxon signed rank test with a test value of 4 was used for all constructs but the exercise assessment scale, where a test value of 3 was used.

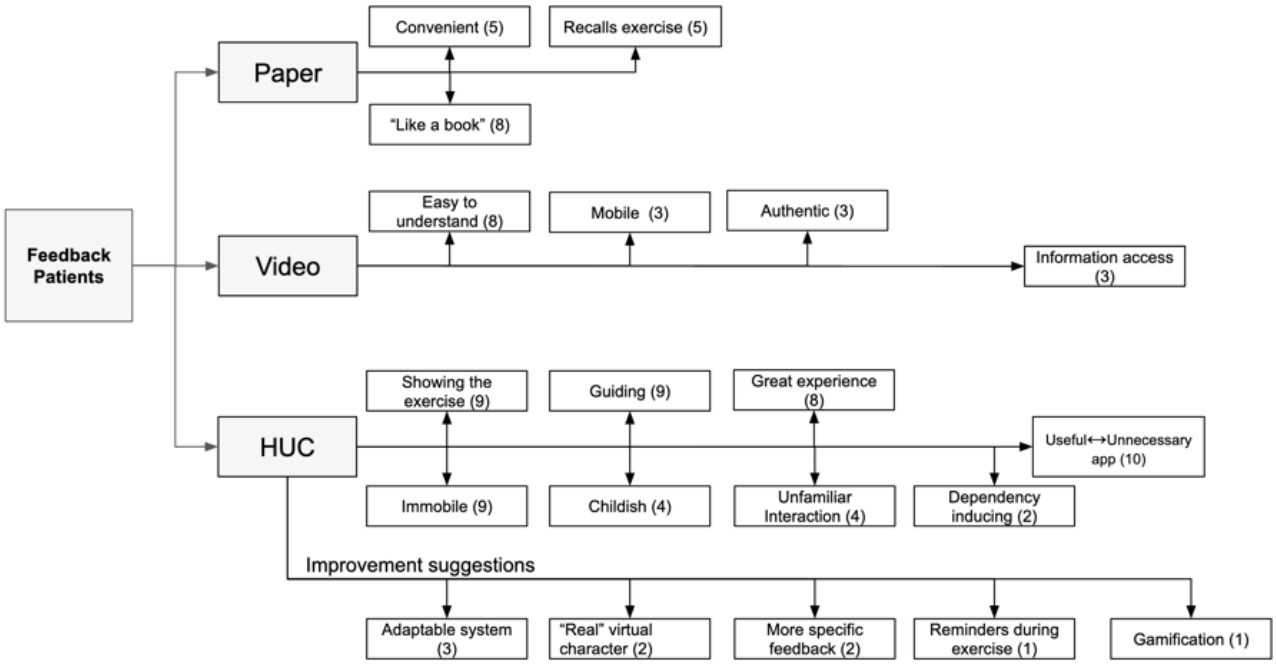
[d]7-point Likert scales ranging from very low (1) to very high (7).

[e]7-point Likert scales ranging from strongly disagree (1) to strongly agree (7).

[f]N/A: not applicable.

[g]Exercise assessment scale items resulted in a score from completely incorrect (0) to completely correct (6).

Figure 9. Thematic map of the patients' qualitative feedback of study 3. Note: the number in brackets indicates the frequency a topic was mentioned. HUC: hybrid ubiquitous coaching.



Study 4: HUC Assessed in a 4-Week Field Study

Frequency and Engagement

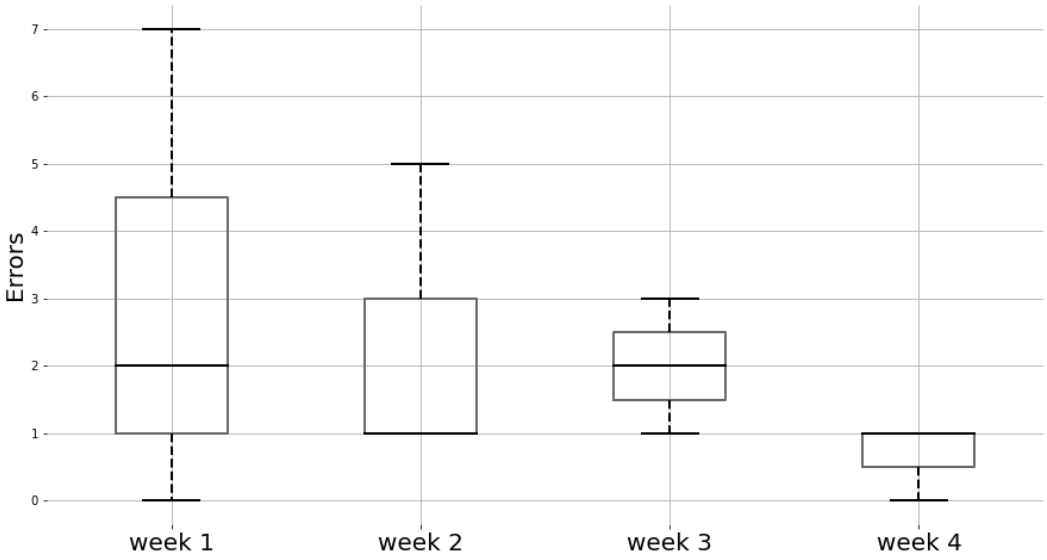
Based on the behavioral data recorded by the HUC, a total of 3 exercise sessions per week were performed in 3 out of 4 weeks. This resulted in TAD rates of 92% in frequency, sets, and repetitions (Table 5). During the last week, the patient was on vacation and did not take the hardware with them. Therefore, only 2 sessions were performed. Not going low enough with the body during the squat exercise was the most common mistake. The average number of mistakes during the 4 weeks is depicted in Figure 10. The number of mistakes fell over the 4 weeks, indicating that the HUC's real-time feedback was processed by the patient and, at least to some degree, put into action. In total, 17 standardized text messages were sent to remind the patient to perform the exercises, to inform the patient about their progress, and to provide motivational information about the importance of performing the exercises. The patient acknowledged these messages with 8 answers (Multimedia Appendix 12).

Table 5. Behavioral data from the 4-week HUC practice (N=1).

Construct	Value
Session completion rate, n (%)	11 (92)
Repetition rate, n (%)	330 (92)
Sets, n (%)	33 (92)
Errors (1st three sets), n (%)	8 (89)
Errors (last three sets), n (%)	4 (44)
HUC[a] messages, mean (sum)	4.25 per week (17 in total)
Patient messages, mean (sum)	2 per week (8 in total)
Duration (s), mean (sum)	191 per session (2098 in total)

[a]HUC: hybrid ubiquitous coaching.

Figure 10. Box plot of the exercise execution errors during the 4 weeks. The number of errors was aggregated for each week.



Debriefing Interview

Overall, the patient described the experience with the HUC as very positive. The exercises guided by the AR-based CA were enjoyable and motivating. The patient liked the feeling of not being alone while performing the exercises and remarked that they preferred the HUC to a human personal trainer, as it was perceived as being "more relaxing." Additionally, the patient pointed out usually having difficulties adhering to home exercises and consequently appreciating the variations of the WhatsApp-based motivational messages from the smartphone-based CA and feedback on the execution accuracy by the AR-based CA, both of which helped the patient with adherence. The patient expressed an interest in having a more dynamic intervention program that includes longer sessions and variations in the exercises. Finally, the patient wanted to continue working with the smartphone-based and AR-based CA beyond the period of this n-of-1 trial.

Follow-up Interviews With Physiotherapists

Reflecting on the results of the study, all physiotherapists confirmed that HUC has the potential to improve adherence to home-based physiotherapy exercises. Referring to the progression of the exercise technique, the real-time feedback about the temporal and spatial accuracy was seen as a clear advantage. Two physiotherapists also liked the fact that the system ensured adherence to the prescribed number of sets and repetitions and described it as having a positive influence on patient motivation. As a treatment provider, the detailed monitoring information regarding the exercise technique was considered to be highly valuable. All physiotherapists highlighted that the therapy could be optimized and a patient's problems could be addressed more individually during on-site therapy sessions. This could potentially also render the face-to-face encounters more efficient. In a future version of HUC, patients should be able to give feedback on their level of pain or well-being. Particularly, information about a patient's pain level could enable more personalized digital coaching. Lastly, according to 2 therapists, it should be possible to calibrate the exercises more individually in terms of exercise technique by adding different variations of one exercise.

Discussion

Principal Findings

The goal of this paper was to address the problem of nonadherence to home exercises by proposing HUC, a human- and CA-supported coaching approach that employs smartphone and AR technology. This paper is, to the best of our knowledge, the first to investigate the potential of HUC with patients and health care experts in two lab studies, interview studies, and a longitudinal n-of-1 trial in the field. HUC required collaborative and interdisciplinary effort from various stakeholders, including health care experts, patients, and experts in behavioral medicine, game design, AR, and human computer interaction. A strength of this work and HUC is, therefore, that its pragmatic and diverse investigation helps to broadcast the results to real-world scenarios and inform practitioners about effective digital designs for addressing the problem of treatment adherence. Moreover, the implemented home exercise (ie, the squat) is not only relevant to physiotherapy patients but also an important component of other treatments, for example, of high-intensity interval trainings for patients who are overweight or patients with cardiovascular disease. In the following section, the results will be discussed along the research questions and design challenges. Limitations and future work will also be outlined.

Research Question 1a: How is HUC Perceived by Patients?

Overall, patients stated that they enjoyed executing the exercise together with HUC, that they intended to use it, and that they would recommend the final HUC version to other patients.

Compared to paper- and video-based instructions, HUC reached higher overall acceptance rates, required less mental capacity, was perceived as being easier to follow, and resulted in higher exercise accuracy ratings. However, significant differences were only found for the items *mental capacity*, *perceived enjoyment*, and *perceived usefulness* between HUC and the paper-based instructions. This is still promising, however, for the following reasons. First, HUC was rated at least as good as video instructions, which is encouraging, since video instructions are associated with reduced symptoms in physiotherapy [103]. Second, perceived enjoyment was significantly higher than with paper-based instructions, a promising result, since enjoyment is related to engagement in digital health interventions [104], something that is often not taken into consideration [105] and is important for long-term adherence. This was confirmed in the 4-week field study. In line with previous findings in the field of CAs [47,49,57,59], patients reported that they were able to build a working alliance with the CA. This was especially the case for attachment bonds, which include factors such as mutual trust, acceptance, and empathy [49,106]. Lastly, in the field study, the patient indicated that they were more motivated to do the exercise when using HUC. The positive perception of HUC and the feeling of collaboration between the patient and the smartphone- and AR-based CA implies an established working alliance.

In summary and to answer RQ1a, we conclude that patients accepted and enjoyed HUC. Patients were also able to build a working alliance with the CA.

Research Question 1b: How is HUC Perceived by Health Care Experts?

HUC was also evaluated by physiotherapists to gather their feedback. Collecting behavioral data about various treatment adherence dimensions (eg, the session completion rate, the sets rate, and the exercise repetition rate) was seen as a clear advantage of HUC. Physiotherapists also considered the real-time feedback to be an additional motivator due to its personal, interactive, and gamified approach, which reflects an improvement relative to the status quo. However, physiotherapists were initially (study 2) more skeptical about whether HUC could be conducive to building a strong working alliance between patients and physiotherapists. The importance of personal interaction between the physiotherapist and the patient was one reason for the negative assessment of attachment bond. Physiotherapists expressed that mutual trust, empathy, and, consequently, mutual goals and tasks were uniquely established during face-to-face encounters. Nevertheless, both the quantitative and qualitative patient feedback (study 3) underline the potential that a strong working alliance can be built with a CA. This result is consistent with prior work about working alliances and CAs [68]. There is a possible explanation for the deviation between the working alliance ratings of physiotherapists (study 2) and patients (study 3): While patients assessed the working alliance with the CA, physiotherapists assessed it in comparison to the bond that develops during physical interaction between patient and therapist. However, HUC foresees the integration of human and CA coaching into both in-person and remote physiotherapy sessions (Figure 1). As HUC does not replace but rather complements the bond between the physiotherapist and patient, it is more equitable to compare HUC with traditional coaching methods. Thus, compared to conventional methods, HUC is shown to have the ability to increase the working alliance.

In the follow-up interviews about the field study, all physiotherapists agreed that home exercises in physiotherapy could be optimized with HUC. They were convinced that various treatment adherence dimensions could be addressed more efficiently with HUC.

In summary and to answer RQ1b, we conclude that HUC represents a clear relative advantage to current methods, as indicated by physiotherapists, in particular, with respect to its motivational features and real-time feedback.

Research Question 2: Does HUC Lead to Treatment adherence to Home Exercises?

Results from study 3 and the 4-week n-of-1 trial were promising with respect to RQ2. Qualitative feedback from study 3 indicated that HUC addressed adherence challenges better than the paper- and video-based instruction methods. In particular, patients found the reminders (TAD1-2) useful, appreciated receiving guidance (TAD3) and real-time feedback on the spatial and temporal accuracy of the exercise execution (TAD4-5), and felt more comfortable about the exercise. Moreover, the results of the 4-week trial revealed a clear decrease in exercise execution errors (TAD4-5) and a high adherence rate of 92% in TAD1-3. Lastly, the patient reported being motivated and committed to doing the exercise doing the 4 weeks due to the variations in the messages and the feedback on exercise accuracy. Motivation and enjoyment are targeted by HUC and, according to prior work, associated with significantly increased rates of adherence [14,29,30].

In summary and to answer RQ2, we conclude that HUC did address the session completion rate (TAD1), set completion rate (TAD2), exercise repetition rate (TAD3), and temporal and spatial exercise accuracy (TAD5). Although this study does not prove beyond a doubt that the use of HUC led to better adherence, the rate was far higher than has typically been found in previous studies [18,20] and was also higher than the participant's previous adherence to home physiotherapy exercise based on their self-report of their own practice and difficulties.

Design Challenges

A first challenge during the design of HUC was the development of the CA's personality in such a way that the patient perceives the CA coherently as the physiotherapist's digital assistant via text messages on the smartphone and voice messages in the AR space. The first lab study and the NPS revealed that HUC, and the design in particular, still needed improvement. Accordingly, the first version of HUC was modified based on this feedback. In the second lab study, patients reported correspondently higher levels of satisfaction with HUC and acceptance of the design of the CA. Further, another challenge was the design of the AR interactions, which had to take into account the limited exercise space at patients' homes, the limited field of view of the AR glasses, and anthropometric characteristics of patients. Any additional controller hardware also had to be omitted due to the requirement that physiotherapy exercises be hands-free. This

challenge was partly met. Patients ranked HUC as their first-choice coaching method but still commented on the bulkiness of the hardware. Another major challenge was the design of variations in the smartphone-based and AR-based CA conversations to improve long-term adherence to physiotherapy in the last study. Variations in the design were implemented to make the exercises more diversified and to thereby increase adherence. The positive result for treatment adherence indicated that the last major design challenge was met.

Limitations and Future Work

The current work has the following limitations. First and foremost, in the current HUC version, only one home exercise was implemented, and thus the findings, especially those related to the AR-based CA, cannot be generalized to a large extent. Implementations of HUC should be further developed to reach a level of implementation foreseen in its approach, as outlined in Figure 1. Future work is therefore required to develop an intuitive CA that supports health care experts through the customization of home exercise sessions, for example, via suggestions given by prior patients [107,108]. Additionally, the problem with the deviations in the exercises could also be solved with machine learning. In doing so, exercise suggestions could be learned, for example, from a basket analysis that considers past exercise programs, patient-related variables (eg, diagnosis and preferences) and past treatment outcomes (eg, reduction in pain) [109,110].

Second, objective adherence rates were based on lab studies and a 4-week study. Regular physical therapy treatment lasts approximately 9 weeks, while long-term treatment can last more than 9 months (36 sessions) [111]. Thus, HUC needs to be assessed not only by a representative sample of patients but also for a representative treatment duration that is common not only in physiotherapy but also in other treatments that require intensive home exercises (eg, treatments for patients with obesity or cardiovascular disease).

Third, the current AR-based CA had a fixed coaching approach. A growing body of research reveals that certain personalities and therapeutic techniques (eg, interpersonal, person-centered, behavioral) work better in specific contexts [112,113]. The suitability of the coaching approach can also depend on individuals' personality traits [114]. By tailoring the coaching technique, the CA could be perceived as more empathetic, potentially making the coaching experience more enjoyable and thereby increasing both adherence and the working alliance and with them, treatment outcomes [112]. Further, we did not investigate whether the prescribed exercises actually had an effect on health outcomes. An adaptive algorithm that increases the difficulty, repetition, and intensity of the exercises could be implemented to ensure that the patient always performs slightly above the comfort zone. This might have an effect not only on health outcomes, but also on adherence. Lastly, the AR-based part of HUC was not yet perceived to be mobile enough, and this was a frequently expressed reason why patients were doubtful about using HUC. Nevertheless, this study was first and foremost about testing real-time feedback and a hands-free interaction paradigm with a "human-sized" CA. The AR hardware is currently changing and shrinking significantly [115], and future work will therefore rely on significantly smaller AR and will be able to test HUC on sufficiently more wearable hardware.

Conclusions

This work provides evidence of the relevance and utility of HUC that aims to increase adherence to home exercises. It therefore contributes to the field of digital health by outlining how CAs hosted by mobile and wearable technologies can extend the reach and effectiveness of health care experts into the everyday lives of patients. HUC may be promising not only in the context of physiotherapy, as exemplarily elaborated in this work, but also for various other conditions, such as obesity or cardiovascular disease, as they also require intensive and longitudinal behavioral support in the everyday lives of patients.

Authors' Contributions

TK, VE, LS, and EMH contributed to the study design. HG provided technical support for the study execution. KML, VE, and TK contributed to the data analysis of the results. All authors provided critical review and revision of the manuscript. All authors approved the manuscript before submission.

Conflicts of Interest

TK, KML, LS, and HG are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Department of Management, Technology and Economics at ETH Zurich and the Institute of Technology Management at the University of St Gallen, which is funded in part by the Swiss health insurer CSS. TK is also a cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, Pathmate Technologies is not involved in any of the 4 studies described in this paper. All other authors declare no conflict of interest.

Multimedia Appendix 1
Current State Commercial and Research Applications.
[DOCX File , 67 KB - jmir_v23i2e23612_app1.docx]

Multimedia Appendix 2
Storybook of the hybrid ubiquitous coaching approach.
[DOCX File , 179 KB - jmir_v23i2e23612_app2.docx]

Multimedia Appendix 3
First-person view and interactions with the augmented reality-based conversational agent version 1 of study 1.
[MP4 File (MP4 Video), 165488 KB - jmir_v23i2e23612_app3.mp4]

Multimedia Appendix 4
Technical details of the augmented reality-based conversational agents versions 1 and 2.
[DOCX File , 15 KB - jmir_v23i2e23612_app4.docx]

Multimedia Appendix 5
First-person view and interactions with the augmented reality-based conversational agent version 2.
[MP4 File (MP4 Video), 197084 KB - jmir_v23i2e23612_app5.mp4]

Multimedia Appendix 6
Sketches of the hybrid ubiquitous coaching approach.
[PDF File (Adobe PDF File), 6067 KB - jmir_v23i2e23612_app6.pdf]

Multimedia Appendix 7
Questions of the semistructured interview of study 2.
[DOCX File , 28 KB - jmir_v23i2e23612_app7.docx]

Multimedia Appendix 8
Video-based instructions of study 3.
[MP4 File (MP4 Video), 20604 KB - jmir_v23i2e23612_app8.mp4]

Multimedia Appendix 9
Quantitative questionnaire items and depersonalized raw data of all 4 studies.
[XLSX File (Microsoft Excel File), 28 KB - jmir_v23i2e23612_app9.xlsx]

Multimedia Appendix 10
Questions of the semistructured interview of study 3.
[DOCX File , 15 KB - jmir_v23i2e23612_app10.docx]

Multimedia Appendix 11
Data analysis script written in Python for all 4 studies.
[ZIP File (Zip Archive), 3 KB - jmir_v23i2e23612_app11.zip]

Multimedia Appendix 12
Examples of the smartphone-based conversational turns of study 4.
[DOCX File , 987 KB - jmir_v23i2e23612_app12.docx]

References

1. Tsang A, Von Korff M, Lee S, Alonso J, Karam E, Angermeyer MC, et al. Common chronic pain conditions in developed and developing countries: gender and age differences and comorbidity with depression-anxiety disorders. J Pain 2008 Oct;9(10):883-891. [doi: 10.1016/j.jpain.2008.05.005] [Medline: 18602869]
2. Briggs AM, Cross MJ, Hoy DG, Sànchez-Riera L, Blyth FM, Woolf AD, et al. Musculoskeletal Health Conditions Represent a Global Threat to Healthy Aging: A Report for the 2015 World Health Organization World Report on Ageing and Health. Gerontologist 2016 Apr;56 Suppl 2:S243-S255. [doi: 10.1093/geront/gnw002] [Medline: 26994264]
3. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
4. Briggs AM, Woolf AD, Dreinhöfer K, Homb N, Hoy DG, Kopansky-Giles D, et al. Reducing the global burden of musculoskeletal conditions. Bull World Health Organ 2018 May 01;96(5):366-368 [FREE Full text] [doi: 10.2471/BLT.17.204891] [Medline: 29875522]

5. March L, Smith EUR, Hoy DG, Cross MJ, Sanchez-Riera L, Blyth F, et al. Burden of disability due to musculoskeletal (MSK) disorders. Best Pract Res Clin Rheumatol 2014 Jun;28(3):353-366. [doi: 10.1016/j.berh.2014.08.002] [Medline: 25481420]
6. Malik KM, Beckerly R, Imani F. Musculoskeletal Disorders a Universal Source of Pain and Disability Misunderstood and Mismanaged: A Critical Analysis Based on the U.S. Model of Care. Anesth Pain Med 2018 Dec;8(6):e85532 [FREE Full text] [doi: 10.5812/aapm.85532] [Medline: 30775292]
7. Friedrich M, Gittler G, Arendasy M, Friedrich KM. Long-term effect of a combined exercise and motivational program on the level of disability of patients with chronic low back pain. Spine (Phila Pa 1976) 2005 May 01;30(9):995-1000. [doi: 10.1097/01.brs.0000160844.71551.af] [Medline: 15864148]
8. Hayden JA, van TMW, Tomlinson G. Systematic review: strategies for using exercise therapy to improve outcomes in chronic low back pain. Ann Intern Med 2005 May 03;142(9):776-785. [Medline: 15867410]
9. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. Man Ther 2010 Jun;15(3):220-228 [FREE Full text] [doi: 10.1016/j.math.2009.12.004] [Medline: 20163979]
10. Shori G, Kapoor G, Talukdar P. Effectiveness of home-based physiotherapy on pain and disability in participants with osteoarthritis of knee: an observational study. J Phys Ther Sci 2018 Oct;30(10):1232-1236 [FREE Full text] [doi: 10.1589/jpts.30.1232] [Medline: 30349155]
11. World Health Organization. Adherence to long-term therapies: evidence for action. 2003. URL: https://apps.who.int/iris/bitstream/handle/10665/42682/9241545992.pdf [accessed 2021-01-18]
12. Campbell R, Evans M, Tucker M, Quilty B, Dieppe P, Donovan JL. Why don't patients do their exercises? Understanding non-compliance with physiotherapy in patients with osteoarthritis of the knee. J Epidemiol Community Health 2001 Feb;55(2):132-138 [FREE Full text] [Medline: 11154253]
13. Friedrich M, Gittler G, Halberstadt Y, Cermak T, Heiller I. Combined exercise and motivation program: effect on the compliance and level of disability of patients with chronic low back pain: a randomized controlled trial. Arch Phys Med Rehabil 1998 May;79(5):475-487. [Medline: 9596385]
14. Beinart NA, Goodchild CE, Weinman JA, Ayis S, Godfrey EL. Individual and intervention-related factors associated with adherence to home exercise in chronic low back pain: a systematic review. Spine J 2013 Dec;13(12):1940-1950. [doi: 10.1016/j.spinee.2013.08.027] [Medline: 24169445]
15. Austin S, Qu H, Shewchuk RM. Association between adherence to physical activity guidelines and health-related quality of life among individuals with physician-diagnosed arthritis. Qual Life Res 2012 Oct;21(8):1347-1357. [doi: 10.1007/s11136-011-0046-x] [Medline: 22038394]
16. Campbell AG, Stafford JW, Holz T, O'Hare GMP. Why, when and how to use augmented reality agents (AuRAs). Virtual Reality 2013 Dec 1;18(2):139-159. [doi: 10.1007/s10055-013-0239-4]
17. Vasey LM. DNAs and DNCTs — Why Do Patients Fail to Begin or to Complete a Course of Physiotherapy Treatment? Physiotherapy 1990 Sep;76(9):575-578. [doi: 10.1016/s0031-9406(10)63052-0]
18. Medina-Mirapeix F, Escolar-Reina P, Gascón-Cánovas JJ, Montilla-Herrador J, Jimeno-Serrano FJ, Collins SM. Predictive factors of adherence to frequency and duration components in home exercise programs for neck and low back pain: an observational study. BMC Musculoskelet Disord 2009 Dec 09;10:155 [FREE Full text] [doi: 10.1186/1471-2474-10-155] [Medline: 19995464]
19. Härkäpää K, Järvikoski A, Mellin G, Hurri H, Luoma J. Health locus of control beliefs and psychological distress as predictors for treatment outcome in low-back pain patients: results of a 3-month follow-up of a controlled intervention study. Pain 1991 Jul;46(1):35-41. [doi: 10.1016/0304-3959(91)90031-R] [Medline: 1832753]
20. Reilly K, Lovejoy B, Williams R, Roth H. Differences between a supervised and independent strength and conditioning program with chronic low back syndromes. J Occup Med 1989 Jun;31(6):547-550. [doi: 10.1097/00043764-198906000-00012] [Medline: 2525182]
21. Kolt GS, McEvoy JF. Adherence to rehabilitation in patients with low back pain. Man Ther 2003 May;8(2):110-116. [Medline: 12890439]
22. McLean SM, Burton M, Bradley L, Littlewood C. Interventions for enhancing adherence with physiotherapy: a systematic review. Man Ther 2010 Dec;15(6):514-521. [doi: 10.1016/j.math.2010.05.012] [Medline: 20630793]
23. Smith M, Davis MA, Stano M, Whedon JM. Aging baby boomers and the rising cost of chronic back pain: secular trend analysis of longitudinal Medical Expenditures Panel Survey data for years 2000 to 2007. J Manipulative Physiol Ther 2013 Jan;36(1):2-11 [FREE Full text] [doi: 10.1016/j.jmpt.2012.12.001] [Medline: 23380209]
24. Frost R, Levati S, McClurg D, Brady M, Williams B. What Adherence Measures Should Be Used in Trials of Home-Based Rehabilitation Interventions? A Systematic Review of the Validity, Reliability, and Acceptability of Measures. Arch Phys Med Rehabil 2017 Jun;98(6):1241-1256.e45. [doi: 10.1016/j.apmr.2016.08.482] [Medline: 27702555]
25. Friedrich M, Cermak T, Maderbacher P. The effect of brochure use versus therapist teaching on patients performing therapeutic exercise and on changes in impairment status. Phys Ther 1996 Oct;76(10):1082-1088. [Medline: 8863761]
26. Bachmann C, Oesch P, Bachmann S. Recommendations for Improving Adherence to Home-Based Exercise: A Systematic Review. Phys Med Rehab Kuror 2017 Dec 18;28(01):20-31. [doi: 10.1055/s-0043-120527]

27. Alexandre NMC, Nordin M, Hiebert R, Campello M. Predictors of compliance with short-term treatment among patients with back pain. Rev Panam Salud Publica 2002 Aug;12(2):86-94. [Medline: 12243693]
28. Iversen MD, Fossel AH, Katz JN. Enhancing function in older adults with chronic low back pain: a pilot study of endurance training. Arch Phys Med Rehabil 2003 Sep;84(9):1324-1331. [doi: 10.1016/s0003-9993(03)00198-9] [Medline: 13680569]
29. Escarti A, Guzman J. Effects of feedback on self-efficacy, performance, and choice in an athletic task. Journal of Applied Sport Psychology 1999 Mar 11;11(1):83-96. [doi: 10.1080/10413209908402952]
30. Martin LR, Williams SL, Haskard KB, Dimatteo MR. The challenge of patient adherence. Ther Clin Risk Manag 2005 Sep;1(3):189-199 [FREE Full text] [Medline: 18360559]
31. Chan DK, Lonsdale C, Ho PY, Yung PS, Chan KM. Patient motivation and adherence to postsurgery rehabilitation exercise recommendations: the influence of physiotherapists' autonomy-supportive behaviors. Arch Phys Med Rehabil 2009 Dec;90(12):1977-1982. [doi: 10.1016/j.apmr.2009.05.024] [Medline: 19969157]
32. Basset S. The assessment of patient adherence to physiotherapy rehabilitation. New Zeal J Physiother 2003;31(2):60-66.
33. Picorelli AMA, Pereira LSM, Pereira DS, Felício D, Sherrington C. Adherence to exercise programs for older people is influenced by program characteristics and personal factors: a systematic review. J Physiother 2014 Sep;60(3):151-156 [FREE Full text] [doi: 10.1016/j.jphys.2014.06.012] [Medline: 25092418]
34. Schoo A, Morris M, Bui Q. The effects of mode of exercise instruction on compliance with a home exercise program in older adults with osteoarthritis. Physiotherapy 2005 Jun;91(2):79-86. [doi: 10.1016/j.physio.2004.09.019]
35. Schneiders A, Zusman M, Singer K. Exercise therapy compliance in acute low back pain patients. Manual Therapy 1998 Aug;3(3):147-152. [doi: 10.1016/S1356-689X(98)80005-2]
36. Chung BPH, Chiang WKH, Lau H, Lau TFO, Lai CWK, Sit CSY, et al. Pilot study on comparisons between the effectiveness of mobile video-guided and paper-based home exercise programs on improving exercise adherence, self-efficacy for exercise and functional outcomes of patients with stroke with 3-month follow-up: A single-blind randomized controlled trial. Hong Kong Physiother J 2020 Jun;40(1):63-73 [FREE Full text] [doi: 10.1142/S1013702520500079] [Medline: 32489241]
37. Banos O, Nugent C. M-Coaching: Towards the Next Generation of Mobile-Driven Healthcare Support Services. Computer 2018 Aug;51(8):14-17. [doi: 10.1109/mc.2018.3191267]
38. Lambert TE, Harvey LA, Avdalis C, Chen LW, Jeyalingam S, Pratt CA, et al. An app with remote support achieves better adherence to home exercise programs than paper handouts in people with musculoskeletal conditions: a randomised trial. J Physiother 2017 Jul;63(3):161-167 [FREE Full text] [doi: 10.1016/j.jphys.2017.05.015] [Medline: 28662834]
39. Verapy Health. Verapy. 2016. URL: https://verapytherapy.com/ [accessed 2021-01-18]
40. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
41. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
42. Healure Technology. Healure: Physiotherapy Exercise Plans. 2017. URL: http://www.healure.com [accessed 2020-01-18]
43. Cassell J. Embodied conversational interface agents. Commun. ACM 2000 Apr;43(4):70-78. [doi: 10.1145/332051.332075]
44. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
45. de Cock C, Milne-Ives M, van Velthoven M, Alturkistani A, Lam C, Meinert E. Effectiveness of Conversational Agents (Virtual Assistants) in Health Care: Protocol for a Systematic Review. JMIR Res Protoc 2020 Mar 9;9(3):e16934. [doi: 10.2196/16934]
46. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
47. Kowatsch T, Nißen M, Shih, Chen-Hsuan I, Rüegger D, Volland D, Filler A, et al. Text-based Healthcare Chatbots Supporting Patient and Health Professional Teams: Preliminary Results of a Randomized Controlled Trial on Childhood Obesity. 2017 Aug Presented at: Embodied Agents for Behavior Change (PEACH2017) Workshop, co-located with the 17th International Conference on Intelligent Virtual Agents (IVA 2017); August 27, 2017; Stockholm, Sweden p. 1-10.
48. Kramer J, Künzler F, Mishra V, Smith SN, Kotz D, Scholz U, et al. Which Components of a Smartphone Walking App Help Users to Reach Personalized Step Goals? Results From an Optimization Trial. Ann Behav Med 2020 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
49. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A Smartphone-Based Health Care Chatbot to Promote Self-Management of Chronic Pain (SELMA): Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Apr 03;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
50. Prvu Bettger J, Green CL, Holmes DN, Chokshi A, Mather RC, Hoch BT, et al. Effects of Virtual Exercise Rehabilitation In-Home Therapy Compared with Traditional Care After Total Knee Arthroplasty: VERITAS, a Randomized Controlled Trial. J Bone Joint Surg Am 2020 Jan 15;102(2):101-109. [doi: 10.2106/JBJS.19.00695] [Medline: 31743238]

51. Vivira Health Lab GmbH. Vivira - Physiotherapy Exercises at Home. 2016. URL: https://www.vivira.com/ [accessed 2021-01-28]
52. Ma T, Chattopadhyay D, Sharifi H. Virtual humans in health-related interventions: A meta-analysis. New York, NY, USA: ACM; 2019 May 04 Presented at: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '19); May 4-9, 2019; Glasgow, UK p. 1-6. [doi: 10.1145/3290607.3312853]
53. Babylon Health. Babylon. 2016. URL: https://www.babylonhealth.com [accessed 2021-01-18]
54. Bickmore TW, Pfeifer LM, Byron D, Forsythe S, Henault LE, Jack BW, et al. Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials. J Health Commun 2010;15 Suppl 2:197-210. [doi: 10.1080/10810730.2010.499991] [Medline: 20845204]
55. GOREHA GmbH. XR Health. 2018. URL: https://www.xr.health/ [accessed 2021-01-18]
56. Nass C, Steuer J, Tauber ER. Computers are social actors. 1994 Presented at: CHI94: ACM Conference on Human Factors in Computer Systems; April 1994; Boston, MA p. 72-78. [doi: 10.1145/259963.260288]
57. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]
58. Shamekhi A, Bickmore T, Lestoquoy A, Gardiner P. Augmenting group medical visits with conversational agents for stress management behavior change. In: Persuasive Technology: Development and Implementation of Personalized Technologies to Change Attitudes and Behaviors. 2017 Presented at: 12th International Conference, PERSUASIVE 2017; April 4-6, 2017; Amsterdam. [doi: 10.1007/978-3-319-55134-0_5]
59. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy (Chic) 2018 Dec;55(4):316-340. [doi: 10.1037/pst0000172] [Medline: 29792475]
60. Cassell J, Sullivan J, Prevost S, editors. Embodied Conversational Agents. Cambridge, MA: MIT Press; 2000.
61. Milgram P, Takemura H, Utsumi A, Kishino F. Augmented reality: a class of displays on the reality-virtuality continuum. 1995 Presented at: Photonics for Industrial Applications; October 31-November 4, 1994; Boston, MA p. 282-292. [doi: 10.1117/12.197321]
62. Bandura A. Self-efficacy mechanism in human agency. Am Psychol 1982;37(2):122.
63. Schwarzer R. Modeling Health Behavior Change: How to Predict and Modify the Adoption and Maintenance of Health Behaviors. Applied Psychology 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
64. Andersson G, Carlbring P, Berger T, Almlöv J, Cuijpers P. What makes Internet therapy work? Cogn Behav Ther 2009;38 Suppl 1:55-60. [doi: 10.1080/16506070902916400] [Medline: 19675956]
65. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
66. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - A systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016 Dec;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
67. Kowatsch T, Nißen M, Rüegger D, Stieger M, Flückiger C, Allemand M, et al. The Impact of Interpersonal Closeness Cues in Text-based Healthcare Chatbots on Attachment Bond and the Desire to Continue Interacting: An Experimental Design. 2018 Jun 23 Presented at: Proc 26th Eur Conf Inf Syst - ECIS; June 23-28 2018; Portsmouth, UK p. A.
68. Bickmore T, Schulman D, Yin L. Maintaining Engagement in Long-term Interventions with Relational Agents. Appl Artif Intell 2010 Jul 01;24(6):648-666 [FREE Full text] [doi: 10.1080/08839514.2010.492259] [Medline: 21318052]
69. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. Journal of Counseling Psychology 1989 Apr;36(2):223-233. [doi: 10.1037/0022-0167.36.2.223]
70. Brintz CE, Miller S, Olmsted KR, Bartoszek M, Cartwright J, Kizakevich PN, et al. Adapting Mindfulness Training for Military Service Members With Chronic Pain. Mil Med 2020 Mar 02;185(3-4):385-393 [FREE Full text] [doi: 10.1093/milmed/usz312] [Medline: 31621856]
71. Du S, Liu W, Cai S, Hu Y, Dong J. The efficacy of e-health in the self-management of chronic low back pain: A meta analysis. Int J Nurs Stud 2020 Jun;106:103507. [doi: 10.1016/j.ijnurstu.2019.103507] [Medline: 32320936]
72. Bardram J, Friday A. Ubiquitous Computing Systems. In: Ubiquitous Computing Fundamentals. London, UK: Chapman and Hall/CRC; 2018:51-108.
73. Weiser M. The Computer for the 21st Century. Sci Am 1991 Sep;265(3):94-104. [doi: 10.1038/scientificamerican0991-94]
74. Castonguay LG, Constantino MJ, Holtforth MG. The working alliance: Where are we and where should we go? Psychotherapy (Chic) 2006;43(3):271-279. [doi: 10.1037/0033-3204.43.3.271] [Medline: 22122096]
75. Flückiger C, Del Re AC, Wampold BE, Symonds D, Horvath AO. How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. J Couns Psychol 2012 Jan;59(1):10-17. [doi: 10.1037/a0025749] [Medline: 21988681]
76. Daft RL, Lengel RH. Organizational Information Requirements, Media Richness and Structural Design. Management Science 1986 May;32(5):554-571. [doi: 10.1287/mnsc.32.5.554]
77. Tseng F, Cheng T, Li K, Teng C. How does media richness contribute to customer loyalty to mobile instant messaging? Internet Research 2017 Jun 05;27(3):520-537. [doi: 10.1108/intr-06-2016-0181]

78. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
79. Horvath AO, Symonds BD. Relation between working alliance and outcome in psychotherapy: A meta-analysis. Journal of Counseling Psychology 1991;38(2):139-149. [doi: 10.1037/0022-0167.38.2.139]
80. ter Stal S, Kramer LL, Tabak M, op den Akker H, Hermens H. Design Features of Embodied Conversational Agents in eHealth: a Literature Review. International Journal of Human-Computer Studies 2020 Jun;138:102409. [doi: 10.1016/j.ijhcs.2020.102409]
81. Seo DW, Lee JY. Direct hand touchable interactions in augmented reality environments for natural and intuitive user experiences. Expert Systems with Applications 2013 Jul;40(9):3784-3793. [doi: 10.1016/j.eswa.2012.12.091]
82. Schreier M. Fallauswahl. In: Mey G, Mruck K, editors. Handbuch Qualitative Forschung in Der Psychologie. Berlin, Germany: Springer; 2010:238-251.
83. Venkatesh, Morris, Davis, Davis. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 2003;27(3):425. [doi: 10.2307/30036540]
84. Venkatesh, Thong, Xu. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]
85. Davis FD, Bagozzi RP, Warshaw PR. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
86. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003 Dec;81(12):46-54, 124. [Medline: 14712543]
87. Kamis, Koufaris, Stern. Using an Attribute-Based Decision Support System for User-Customized Products Online: An Experimental Investigation. MIS Quarterly 2008;32(1):159. [doi: 10.2307/25148832]
88. Weller SC, Vickers B, Bernard HR, Blackburn AM, Borgatti S, Gravlee CC, et al. Open-ended interview questions and saturation. PLoS One 2018;13(6):e0198606. [doi: 10.1371/journal.pone.0198606] [Medline: 29924873]
89. Flick U. An Introduction To Qualitative Research, 6th ed. London: SAGE Publications; 2018.
90. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res 2007 Aug;42(4):1758-1772 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00684.x] [Medline: 17286625]
91. Moore GC, Benbasat I. Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. Information Systems Research 1991 Sep;2(3):192-222. [doi: 10.1287/isre.2.3.192]
92. Falkenström F, Hatcher RL, Skjulsvik T, Larsson MH, Holmqvist R. Development and validation of a 6-item working alliance questionnaire for repeated administrations during psychotherapy. Psychol Assess 2015 Mar;27(1):169-183. [doi: 10.1037/pas0000038] [Medline: 25346997]
93. Armitage P, Hills M. The Two-Period Crossover Trial. The Statistician 1982 Jun;31(2):119. [doi: 10.2307/2987883]
94. Hart S, Staveland L. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in Psychology 1988;52:139-183. [doi: 10.1016/s0166-4115(08)62386-9]
95. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
96. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a Relational Agent by Hospital Patients with Depressive Symptoms. Interact Comput 2010 Jul 1;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
97. Smith J, Lewis J, Prichard D. Physiotherapy exercise programmes: are instructional exercise sheets effective? Physiother Theory Pract 2005;21(2):93-102. [Medline: 16392462]
98. Dix A. Human–computer interaction: A stable discipline, a nascent science, and the growth of the long tail. Interacting with Computers 2010 Jan;22(1):13-27. [doi: 10.1016/j.intcom.2009.11.007]
99. Razak FHA, Dix A. Doing a Single Person Study in Human Computer Interaction (HCI). Malaysian Journal of Computing 2010;1(1):40-61.
100. Parmar D, Olafsson S, Utami D, Bickmore T. Looking the Part: The Effect of Attire and Setting on Perceptions of a Virtual Health Counselor. New York, NY, USA: ACM; 2018 Nov Presented at: IVA '18: Proceedings of the 18th International Conference on Intelligent Virtual Agents; November; Sydney, Australia p. 301-306. [doi: 10.1145/3267851.3267915]
101. Creed C, Beale R, Cowan B. The Impact of an Embodied Agent's Emotional Expressions Over Multiple Interactions. Interacting with Computers 2014 Jan 29;27(2):172-188. [doi: 10.1093/iwc/iwt064]
102. Riek L. Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. J Human-Robot Interact 2012 Aug 01;1(1):119-136. [doi: 10.5898/jhri.1.1.riek]
103. Huber S, Priebe JA, Baumann K, Plidschun A, Schiessl C, Tölle TR. Treatment of Low Back Pain with a Digital Multidisciplinary Pain Treatment App: Short-Term Results. JMIR Rehabil Assist Technol 2017 Dec 04;4(2):e11 [FREE Full text] [doi: 10.2196/rehab.9032] [Medline: 29203460]
104. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]

105. Micksche M, Sonnbichler G, Isak K. Krebshilfe und Rehabilitation. In: Crevenna R, editor. Onkologische Rehabilitation Grundlagen, Methoden, Verfahren und Wiedereingliederung. Berlin, Germany: Springer; 2020:295-306.
106. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy: Theory, Research & Practice 1979;16(3):252-260. [doi: 10.1037/h0085885]
107. Alves T, Natálio J, Henriques-Calado J, Gama S. Incorporating personality in user interface design: A review. Personality and Individual Differences 2020 Mar;155:109709. [doi: 10.1016/j.paid.2019.109709]
108. Sommers J, Engelbert RHH, Dettling-Ihnenfeldt D, Gosselink R, Spronk PE, Nollet F, et al. Physiotherapy in the intensive care unit: an evidence-based, expert driven, practical statement and rehabilitation recommendations. Clin Rehabil 2015 Nov;29(11):1051-1063 [FREE Full text] [doi: 10.1177/0269215514567156] [Medline: 25681407]
109. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. New York, NY, USA: ACM; 1993 Presented at: Proceedings of the 1993 ACM SIGMOD international conference on Management of data (SIGMOD '93); May 26-28, 1993; Washington, DC, USA p. 207-216.
110. Delgadillo J, Rubel J, Barkham M. Towards personalized allocation of patients to therapists. J Consult Clin Psychol 2020 Sep;88(9):799-808. [doi: 10.1037/ccp0000507] [Medline: 32378910]
111. Halfon P, Eggli Y, Morel Y, Taffé P. The effect of patient, provider and financing regulations on the intensity of ambulatory physical therapy episodes: a multilevel analysis based on routinely available data. BMC Health Serv Res 2015 Feb 07;15:52 [FREE Full text] [doi: 10.1186/s12913-015-0686-6] [Medline: 25889368]
112. Chen R, Rafaeli E, Ziv-Beiman S, Bar-Kalifa E, Solomonov N, Barber JP, et al. Therapeutic technique diversity is linked to quality of working alliance and client functioning following alliance ruptures. J Consult Clin Psychol 2020 Sep;88(9):844-858. [doi: 10.1037/ccp0000490] [Medline: 32584116]
113. Fadhil A, Schiavo G. Designing for Health Chatbots. ArXiv. Preprint posted online on February 24, 2019 [FREE Full text]
114. Yorita A, Egerton S, Oakman J, Chan C, Kubota N. Self-Adapting Chatbot Personalities for Better Peer Support. : IEEE; 2019 Presented at: 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC); 2019; Bari, Italy p. 4094-4100. [doi: 10.1109/smc.2019.8914583]
115. Wilkins M. Optical headwear for augmented reality. United States BAE SYSTEMS plc (London, GB) D865042. 2019 Oct 29. URL: https://www.freepatentsonline.com/D865042.html [accessed 2021-01-18]

Abbreviations

AR: augmented reality
BCT: behavioral change technique
CA: conversational agent
HUC: hybrid ubiquitous coaching
MSD: musculoskeletal disorder
NPS: Net Promoter Score
QR: quick response
RQ: research question
TAD: treatment adherence dimension

Edited by G Eysenbach; submitted 25.08.20; peer-reviewed by A Lau, J Salisbury; comments to author 16.09.20; revised version received 28.09.20; accepted 18.01.21; published 22.02.21.

Please cite as:
Kowatsch T, Lohse KM, Erb V, Schittenhelm L, Galliker H, Lehner R, Huang EM
Hybrid Ubiquitous Coaching With a Novel Combination of Mobile and Holographic Conversational Agents Targeting Adherence to Home Exercises: Four Design and Evaluation Studies
J Med Internet Res 2021;23(2):e23612
URL: https://www.jmir.org/2021/2/e23612
doi:10.2196/23612
PMID:33461957

Original Paper

Effect of a Parent-Focused eHealth Intervention on Children's Fruit, Vegetable, and Discretionary Food Intake (Food4toddlers): Randomized Controlled Trial

Margrethe Røed[1], MEd; Anine C Medin[1], PhD; Frøydis N Vik[1], PhD; Elisabet R Hillesund[1], PhD; Wendy Van Lippevelde[1,2], PhD; Karen Campbell[3], PhD; Nina C Øverby[1], PhD

[1]Department of Nutrition and Public Health, University of Agder, Kristiansand, Norway

[2]Department of Marketing, Innovation and Organisation, Ghent University, Ghent, Belgium

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

Corresponding Author:
Margrethe Røed, MEd
Department of Nutrition and Public Health
University of Agder
Boks 422
Kristiansand, 4604
Norway
Phone: 47 91311933
Email: margrethe.roed@uia.no

Abstract

Background: In Western countries, children's diets are often low in fruits and vegetables and high in discretionary foods. Diet in early life tends to track through childhood and youth and even into adulthood. Interventions should, therefore, be delivered in periods when habitual traits are established, as in toddlerhood when children adapt to their family's diet.

Objective: In this study, we assessed the effect of the Food4toddlers eHealth intervention, which aimed to enhance toddlers' diets by shaping their food and eating environment.

Methods: The Food4toddlers randomized controlled trial was conducted in Norway in 2017-2018. Parent-child dyads were recruited through social media. In total, 298 parents completed an online questionnaire at baseline (mean child age 10.9 months, SD 1.2). Postintervention questionnaires were completed immediately after the intervention (ie, follow-up 1; mean child age 17.8 months, SD 1.3) and 6 months after the intervention (ie, follow-up 2; mean child age 24.2 months, SD 1.9). The intervention was guided by social cognitive theory, which targets the linked relationship between the person, the behavior, and the environment. The intervention group (148/298, 49.7%) got access to the Food4toddlers website for 6 months from baseline. The website included information on diet and on how to create a healthy food and eating environment as well as activities, recipes, and collaboration opportunities. To assess intervention effects on child diet from baseline to follow-up 1 and from baseline to follow-up 2, we used generalized estimating equations and a time × group interaction term. Between-group differences in changes over time for frequency and variety of fruits and vegetables and frequency of discretionary foods were assessed.

Results: At follow-up 1, a significant time × group interaction was observed for the frequency of vegetable intake (P=.02). The difference between groups in the change from baseline to follow-up 1 was 0.46 vegetable items per day (95% CI 0.06-0.86) in favor of the intervention group. No other significant between-group differences in dietary changes from baseline to follow-up 1 or follow-up 2 were observed. However, there is a clear time trend showing that the intake of discretionary foods increases by time from less than 1 item per week at baseline to more than 4 items per week at 2 years of age (P<.001), regardless of group.

Conclusions: A positive intervention effect was observed for the frequency of vegetable intake at follow-up 1 but not at follow-up 2. No other between-group effects on diet were observed. eHealth interventions of longer duration, including reminders after the main content of the intervention has been delivered, may be needed to obtain long-terms effects, along with tailoring in a digital or a personal form.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 92980420; https://doi.org/10.1186/ISRCTN92980420

(J Med Internet Res 2021;23(2):e18311) doi:10.2196/18311

KEYWORDS

toddler; child; eHealth; intervention; randomized controlled trial; fruit; vegetable; discretionary food

Introduction

What toddlers eat is crucial for their health and growth, and in several western countries, young children do not meet dietary guidelines. A specific challenge is the low intake of fruits and vegetables and the high intake of sugar-sweetened beverages and snacks [1-4]. An unhealthy diet early in life increases the risk for overweight, noncommunicable diseases, and certain cancers [5,6].

The World Health Organization's (WHO) report on ending childhood obesity [5] recommends that appropriate and context-specific nutrition information should be easily available for specific target groups and be delivered in ways that are perceived as meaningful for the users. WHO argues that such information is specifically relevant for parents of infants and toddlers. Diet in early life tends to track through childhood and youth and even into adulthood [5,7]. Interventions should, therefore, be delivered when healthy habitual traits are established in the early years, and one of these periods involves the transition from specific baby foods to eating family meals [7-10].

Parents are the primary gatekeepers of child diet in this period [11,12]. To date, few studies assessing the effect of dietary interventions targeting young children through their parents have been undertaken [13,14]. The internet is a popular source for health information among parents, and parents have reported a need for trustworthy, evidence-based, and highly accessible information sources [15-18]. Theory- and evidence-based eHealth interventions, where intervention messages are delivered to the target audience via electronic means and are easily available and accessible for the parents, may fill this information gap. eHealth interventions have the potential to reach many, can easily be changed and adapted to new groups, are available 24/7, and are cost-effective [19-21].

Parental-focused interventions with an emphasis on creating a healthy food and eating environment for the child are recommended and have shown promising results [22,23]. A healthy food environment is characterized by the accessibility and availability of healthy foods for the child and restricted access to unhealthy alternatives [22,24]. In order to create a healthy eating environment, it is essential to incorporate health-promoting feeding practices, such as healthy modeling and repeated exposure to healthy foods [7].

The aim of this study was to examine the effect of a parent-focused eHealth intervention on the child's diet assessed at two time points postintervention. We hypothesized that, compared with the control group, the children in the intervention group would develop a more frequent and varied intake of fruits and vegetables and less frequent intake of discretionary foods from baseline to postintervention.

Methods

Design and Study Population

This study used data from the Food4toddlers randomized controlled trial (RCT), an eHealth intervention aiming to promote a healthy food and eating environment for toddlers. Details of the intervention's design and components have been previously published [25]. The study was a 2-armed RCT involving 298 parent-child dyads. This eHealth intervention was conducted in Norway in 2017-2018. Data were collected at baseline, after 6 months (ie, follow-up 1: postintervention), and after 12 months (ie, follow-up 2: 6 months postintervention). Parents in the intervention group were provided with access to the Food4toddlers website for a period of 6 months after completing the baseline questionnaires. Log-in instructions for the website were sent by email, and up to three reminders were sent to nonresponders. Informed consent from the parents was obtained when they signed in online for participation in the study. A completed CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth) V1.6.1 checklist is available in Multimedia Appendix 1. Research clearance was obtained from the Norwegian Centre for Research Data on June 8, 2016 (reference No. 48643).

Between August 2017 and January 2018, 404 parents of infants and toddlers from across Norway were recruited through a tailored advertisement on Facebook and accepted to participate by signing in at the study home page [26]. Parents of children born between June 2016 and May 2017 were eligible for participation if they were literate in Norwegian. In the case of twins, the parent reported on behalf of the oldest child. All sociodemographic and behavioral data were collected at baseline and follow-up time points using the online survey software SurveyXact (Rambøll) [27]. Up to three email reminders were sent in the absence of a response. Survey items completed by parents concerning children's food intake, parental feeding practices, and demographic data were included in the analysis.

Participants were randomized and allocated to either an intervention or a control group after the baseline data collection. A randomization list was generated in SPSS Statistics for Windows, version 25.0 (IBM Corp), by one of the researchers (NØ) and implemented by the first author (MR). The follow-up questionnaires replicated the baseline questionnaires but also included questions on intervention website use; only the intervention group completed these questions at follow-up 1.

Intervention Development

This Food4toddlers study was developed using the basic steps from the Model of Planned Promotion for Population Health [28]. This intervention is in line with social cognitive theory, which targets the linked relationship between the person, the behavior, and the environment [29] with an emphasis on how to promote action rather than motivation only [28]. The

participants were encouraged to have core foods available, especially vegetables and fruit, both in their home and on the child's plate. The opposite was encouraged for discretionary (ie, noncore) foods and beverages. Food4toddlers was developed in a cocreation process with health care nurses, parents of toddlers, and students and staff at the University of Agder, Norway. Key elements in this process included several individual and group interviews with stakeholders and the inclusion of students in developing and pilot-testing the website. The Food4toddlers eHealth intervention included a website with four main elements: (1) modules, including two to four lessons, covering an introduction and seven topics on promoting healthy food and eating environments for the child; (2) recipes; (3) a discussion forum; and (4) highlighted information about food and beverages, called *Good to know*. In addition, when accessing the Food4toddlers website, a video appeared with information about the study and its focus on how important just a small weekly increase in vegetable consumption may be for the child's long-term health. Small behavioral changes were highlighted with the aim of making the messages easier for the parent to accomplish [30]. The modules had activity elements, such as a quiz or a game, and visual elements that supported the information. During the intervention period, the participants received weekly emails, each containing a link to a new lesson (20 times), thus expanding the content of the intervention. Some lessons were more comprehensive than others, but the estimated time to complete an average lesson was 10 minutes. The Food4toddlers website was available on smartphones and other tablets in the form of a mobile app, in addition to computers.

Measures and Outcomes

Overview

The importance of fruit and vegetable intake for lifelong health is well-documented [3,4], and a diet rich in fruits and vegetables and limited in discretionary foods is the cornerstone of a high-quality diet [2,31-34]. We wanted to measure the frequency of these foods along with the variety of fruits and vegetables, which is less frequently measured [35] and has been shown to be an indicator of preschoolers' overall diet quality [36]. Our previous research using baseline data from the Food4toddlers intervention revealed different patterns in fruit versus vegetable consumption [37]. Therefore, we wanted to further examine this distinction for the intervention's effect and examine the intake of discretionary foods to elaborate on both core and noncore dietary effects. We constructed three separate scales to assess the consumption frequency of vegetables, fruits, and discretionary foods, respectively. Food variety scale scores were calculated separately for vegetables and fruits.

Child Food Intake

Child food intake in this study was assessed by a food frequency questionnaire (FFQ), based on questionnaires previously used in the population-based Norwegian Mother, Father and Child Cohort Study [38] and the nationwide Norwegian diet survey among 12-month-old children [39]. These questionnaires have been previously validated in toddlers [40,41]. Using both questionnaires, we were able to cover a more extensive selection of foods, but different scales made the comparison more difficult. Of the 59 FFQ items in the questionnaire, we used three food groups, comprising 33 items in total.

Assessment of Fruit and Vegetable Intake

Questions covering the intake of fruits and vegetables commonly consumed in Norway [42] included the following: "How often does your child eat the following fruits/vegetables nowadays?" The food items presented included fresh, cooked, or squeezed fruits and vegetables and both homemade and commercially produced variants. In total, 13 vegetables (ie, *carrot, rutabaga, sweet potato, cauliflower, broccoli, green salad, spinach, cucumber, tomato, corn, sweet pepper, pea,* and *other*) and 11 fruits (ie, *orange, banana, apple, pear, plum, grapes, kiwi, melon, mango, berries,* and *other*) were listed.

A 6-point scale, ranging from *never* to *several times a day*, was used with the following response options and recoded into times per week: never or less than once a week = 0, one to three times a week = 2, four to six times a week = 5, once a day = 7, two times a day = 14, and three times or more per day = 24.5. Similar recoding has been done by others [43-46]. We calculated a combined score of total vegetable intake and another for total fruit intake (ie, frequency per day).

The same items, as previously described for the frequency of vegetables and fruits, were used to calculate variety scores of eaten (coded 1) and not eaten (coded 0) vegetables (13 items) and fruits (11 items).

Assessment of Discretionary Foods and Beverages

The questions on the consumption frequency of discretionary foods included the following: "How often does your child eat the following foods nowadays?" The following food groups were assessed: (1) cakes, waffles, and sweet biscuits; (2) desserts and ice cream; (3) chocolate; (4) candy and such; and (5) chips. A 6-point scale was used, ranging from *never* to *several times a day*. The response options were recoded into times per week: never = 0, less than once a week = 0.5, one to three times a week = 2, four to six times a week = 5, one to two times a day = 10.5, and three times or more per day = 24.5.

Beverage intake was assessed with the following question: "How often does your child drink the following drinks nowadays?" Two sugar-sweetened beverages were included. The response options were recoded into daily intake: never/seldom = 0, one to three times a week = 0.29, four to six times a week = 0.71, one per day = 1, two per day = 2, three per day = 3, four per day = 4, and five or more per day = 6. They were then coded into times per week (ie, multiplied by 7) to be consistent with the snack score. Subsequently, we calculated the sum of the combined frequency of intake of discretionary foods per week, including five snack items and two beverage items.

Assessment of Demographics and Use of the Website

Parents reported the following at baseline: child's gender, child's date of birth, whether they lived together with the child's other parent, and their own height, weight, date of birth, and education level. The parent's BMI was calculated from self-reported height and weight (kg/m^2). The categories for parental education level were as follows: primary school or less; primary school plus

one year of, for example, folk high school; high school; vocational school; upper secondary school or less; college or university (≤4 years); college or university (>4 years); other; and don't know. These categories are similar to categories used by others in Norway [39]. The education level was dichotomized: none or up to 4 years of higher-level education and more than 4 years of higher-level education. This cutoff was used since the groups without college or university education were very small (total 11.3%) and since we know that a healthy lifestyle increases for every year of education [47].

From the website, we registered the number of lessons (ie, 22 in total) that the participants in the intervention group had completed. The lessons comprised two to four pages and all of them had to be visited for a lesson to be registered as completed. Lesson number 7 had an element that was only available via a computer; all other lessons were available on different devices.

Statistics

The sample size was calculated for one of the primary outcomes: child diet quality. Because no data on healthy eating scores for Norwegian toddlers are available, the calculation for this study was based on the study of Angelopoulos et al [48]. They used a healthy diet score of 10 components to assess child diet and observed a mean score of 60.5 (SD 9.2). A 3-point difference in score between the control and intervention groups was considered relevant from a public health perspective. From this, we estimated that 142 children in each group would be required to demonstrate statistical significance with a statistical power of 80% and α level of 5%. Assuming a 40% loss to follow-up, we aimed to recruit 237 parents in each group.

Means with standard deviations for continuous variables and frequencies and percentages for categorical variables were reported for baseline characteristics.

We used generalized estimating equations (GEEs) to determine whether the intervention had an effect on child diet from baseline to follow-up 1 and from baseline to follow-up 2. GEEs are suited for identifying how much a sample's average response changes with a one-unit increase in a covariant, which means that all respondents can be included in the analyses even though there are missing responses on the follow-up questionnaires [49]. This method also takes into account the problem with individual correlated data [49]. Frequency of intake (ie, vegetables, fruits, and discretionary foods) and variety of intake (ie, vegetables and fruits) were included as dependent variables in separate models. An interaction term between group (ie, intervention vs control) and time (ie, baseline vs postintervention) was entered into all models to examine the possible effects of the intervention. Specifically, we investigated whether changes in dietary intake from baseline to postintervention periods (ie, follow-up 1 and follow-up 2) differed significantly between the control and intervention groups. An unstructured covariance matrix and robust estimates of the standard error were used. All models were adjusted for child gender and age as well as for parental BMI, education level, and age reported at baseline. We selected covariates based on previous research on determinants for vegetable and fruit intake [50] and in line with the protocol for the study [25]. We ran *t* tests and Mann-Whitney *U* tests as sensitivity analyses, using complete cases and the difference between baseline and follow-up 1 values and baseline and follow-up 2 values for all outcome variables. The intention - to - treat principle was used in the analyses [51]. All analyses were conducted in SPSS Statistics for Windows, version 25.0 (IBM Corp), except for GEEs, which were run in Stata, version 16 (StataCorp LLC). Statistical significance level was set at $P \leq .05$.

Availability of Data and Materials

The data set supporting the conclusions of this article will be available in the University of Agder Open Research repository [52].

Ethics Approval, Trial Registration, and Consent to Participate

This trial was approved by the Norwegian Centre for Research Data on June 29, 2017 (reference No. 48643). This trial was registered at the International Standard Randomised Controlled Trial Number (ISRCTN) registry on September 13, 2017 (trial No. 92980420). Written consent was obtained from all parents on the study home page [26] when they chose to sign up for participation.

Results

Characteristics of the Study Sample

Figure 1 shows the flow of participants in the study. Of the 404 parents that signed up to participate, 298 (73.8%) completed the baseline questionnaire and were included in the study. After the baseline data collection, parents were randomized into the intervention group (148/298, 49.7%) or the control group (150/298, 50.3%). In total, 1 child was erroneously included in the study because he or she was too young and 6 participants had missing data on demographic variables at baseline (ie, parental age, BMI, or education level); they were excluded from the analyses in this paper. Table 1 shows group comparisons of baseline characteristics and food intake measures between participants in the intervention and control groups. There were no significant differences between the groups. Tables 2 and 3 show group comparisons of baseline characteristics between participants retained in this study at follow-up 1 (ie, immediately after the intervention) and follow-up 2 (ie, 6 months postintervention), respectively, and those who were lost to follow-up or had missing data on outcome variables at these time points.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram for the Food4toddlers randomized controlled trial study.

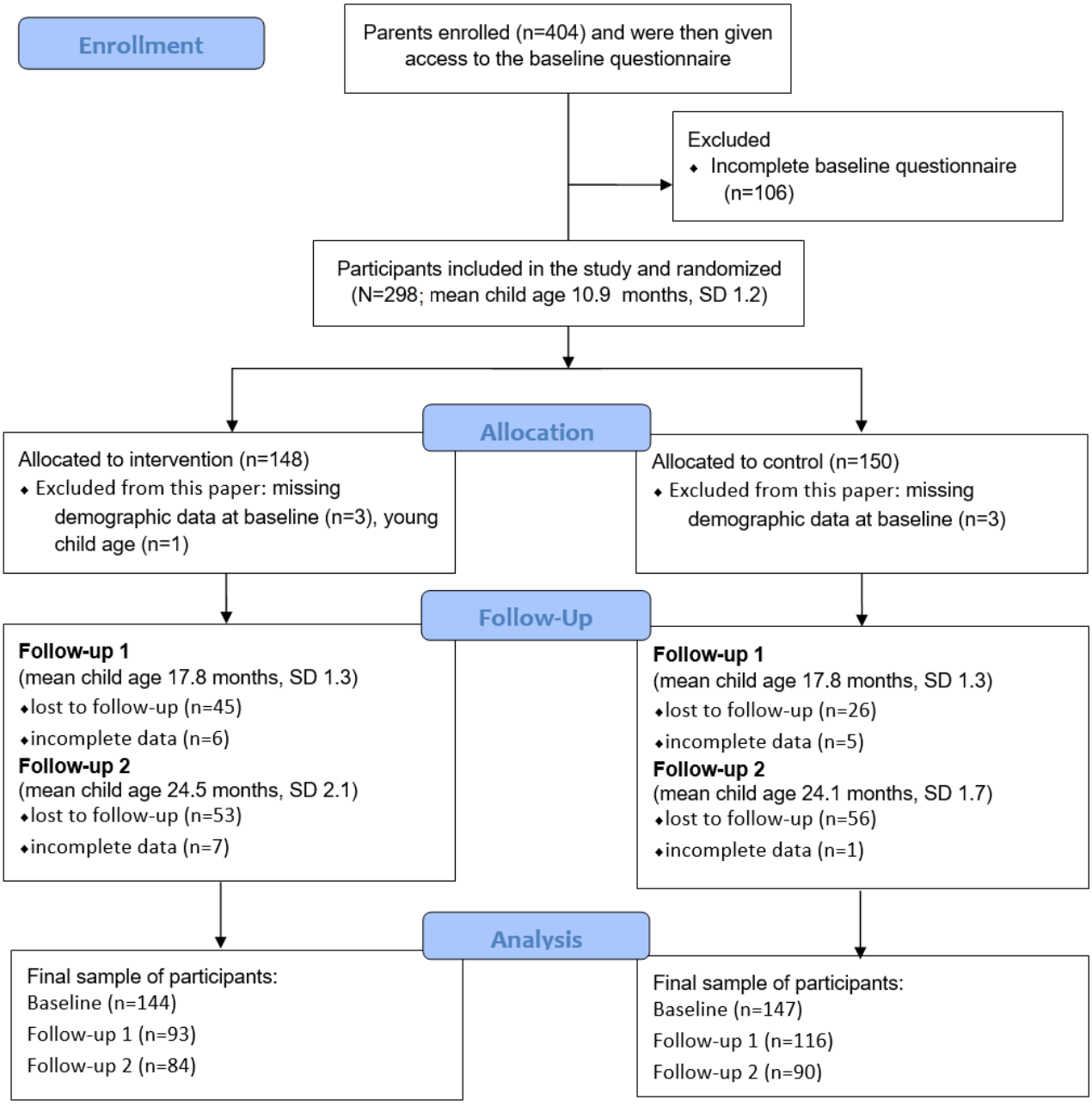


Table 1. Baseline characteristics of parents and children in the intervention and control groups and children's food intake measures.

Characteristics and food intake measures	Total (n=291)	Control (n=147)	Intervention (n=144)
Demographic characteristics			
Parents			
Mother[a], n (%)	287 (98.6)	147 (100)	140 (97.2)
Father, n (%)	4 (1.4)	0 (0)	4 (2.8)
Age (years), mean (SD)	31.7 (4.2)	31.8 (3.9)	31.5 (4.4)
Height (cm), mean (SD)	168.4 (5.9)	168.1 (5.9)	168.7 (6.0)
Weight (kg), mean (SD)	70.7 (14.2)	71.0 (14.8)	70.4 (13.7)
BMI (kg/m^2), mean (SD)	24.9 (4.6)	25.1 (4.8)	24.7 (4.4)
Two-parent household, n (%)[b]	288 (99.0)	144 (98.0)	144 (100)
Total number of household members, mean (SD)	3.6 (0.9)	3.7 (0.9)	3.6 (1.0)
Born in Norway, n (%)	250 (86.2)[c]	122 (83.0)[c]	128 (88.9)
Education: ≤4 years college or university or lower, n (%)	132 (45.4)	70 (47.6)	62 (43.1)
Children			
Age (months), mean (SD)	10.9 (1.2)	10.8 (1.2)	10.9 (1.2)
Child's sex: female, n (%)	129 (44.3)	63 (42.9)	66 (45.8)
Children's food intake at baseline, mean (SD)			
Vegetables (13 items)			
Frequency (times/day)	3.2 (1.6)	3.1 (1.5)	3.2 (1.7)
Variation (number/week)	7.2 (2.6)	7.2 (2.5)	7.2 (2.7)
Fruits (11 items)			
Frequency (times/day)	2.8 (1.6)	2.7 (1.4)	2.9 (1.8)
Variation (number/week)	5.8 (2.2)	5.7 (2.2)	5.9 (2.2)
Discretionary foods (7 items)[d]			
Frequency (times/week)	0.8 (1.4)	0.8 (1.4)	0.8 (1.4)

[a]Included co-mothers and foster mothers.

[b]Lived together with the other parent.

[c]One missing case.

[d]Included five unhealthy snack items and two sugar-sweetened beverages.

Table 2. Differences in baseline characteristics and food intake between participants who remained in the study and those lost to follow-up 1.

Demographic characteristics and food intake measures	At follow-up 1 (all participants) (n=291)			Lost to follow-up 1 from intervention and control groups[a] (n=82)		
	Retained in study (n=209)	Lost to follow-up 1[a] (n=82)	*P* value[b]	Control (n=31)	Intervention (n=51)	*P* value[b]
Demographic characteristics						
Parents						
Mother[c], n (%)	206 (98.6)	81 (99)	N/A[d]	31 (100)	50 (98)	N/A
Father, n (%)	3 (1.4)	1 (1)	N/A	0 (0)	1 (2)	N/A
Age (years), mean (SD)	31.8 (4.1)	31.5 (4.4)	.66	32.6 (4.2)	30.9 (4.4)	.09
Height (cm), mean (SD)	168 (5.8)	168 (6.0)	.74	168 (5.9)	169 (6.1)	.48
Weight (kg), mean (SD)	71.0 (15.0)	69.8 (12.0)	.52	67.6 (11.6)	71.2 (12.2)	.19
BMI (kg/m^2), mean (SD)	24.9 (4.9)	24.6 (3.8)	.55	24.0 (3.6)	25.0 (3.9)	.24
Two-parent household, n (%)[e]	207 (99.0)	81 (99)	.82	30 (99)	51 (100)	.20
Total number of household members, mean (SD)	3.6 (1.0)	3.7 (0.8)	.37	3.9 (0.8)	3.6 (0.8)	.08
Born in Norway, n (%)	177 (84.7)	73 (90)[f]	.23	26 (87)[f]	47 (92)	.42
Education: ≤4 years of college or university or lower, n (%)	93 (44.5)	39 (48)	.64	15 (48)	24 (47)	.91
Children						
Age (months), mean (SD)	10.8 (1.2)	11.0 (1.4)	.13	10.8 (1.7)	11.2 (1.2)	.24
Child's sex: female, n (%)	90 (43.1)	39 (48)	.49	12 (39)	27 (53)	.21
Children's food intake at baseline, mean (SD)						
Vegetables (13 items)						
Frequency (times/day)	3.2 (1.6)	3.0 (1.6)	.36	2.9 (1.5)	3.1 (1.7)	.28
Variation (number/week)	7.4 (2.5)	6.9 (2.7)	.20	6.7 (2.6)	7.0 (2.8)	.64
Fruits (11 items)						
Frequency (times/day)	2.9 (1.6)	2.7 (1.6)	.43	2.9 (1.7)	2.6 (1.7)	.48
Variation (number/week)	5.9 (2.2)	5.6 (2.2)	.39	5.6 (2.0)	5.7 (2.4)	.66
Discretionary foods (7 items)[g]						
Frequency (times/day)	0.8 (1.4)	0.9 (1.4)	.48	0.8 (1.2)	1.0 (1.5)	.78

[a]Participants who were lost to follow-up or had incomplete outcome data at follow-up 1.

[b]Calculated by the Pearson chi-square test or *t* test.

[c]Included co-mothers and foster mothers.

[d]N/A: not applicable; it was not relevant to the study to calculate *P* values for gender items.

[e]Lived together with the other parent.

[f]One missing case.

[g]Included five unhealthy snack items and two sugar-sweetened beverages.

Table 3. Differences in baseline characteristics and food intake between participants who remained in the study and those lost to follow-up 2.

Demographic characteristics and food intake measures	At follow-up 2 (all participants) (n=291)			Lost to follow-up 2 from intervention and control groups[a] (n=117)		
	Retained in study (n=174)	Lost to follow-up 2[a] (n=117)	*P* value[b]	Control (n=57)	Intervention (n=60)	*P* value[b]
Demographic characteristics						
Parents						
Mother[c], n (%)	171 (98.3)	116 (99.1)	N/A[d]	57 (100)	59 (98)	N/A
Father, n (%)	3 (1.7)	1 (0.9)	N/A	0 (0)	1 (2)	N/A
Age (years), mean (SD)	32.0 (4.0)	31.2 (4.4)	.10	32.0 (4.1)	30.5 (4.6)	.07
Height (cm), mean (SD)	169 (5.8)	168 (6.0)	.05	168 (6.2)	168 (5.8)	.89
Weight (kg), mean (SD)	71.3 (14.7)	69.8 (13.4)	.37	69.1 (13.6)	70.6 (13.4)	.53
BMI (kg/m^2), mean (SD)	25.0 (4.9)	24.8 (4.2)	.76	24.4 (3.9)	25.1 (4.5)	.37
Two-parent household, n (%)[e]	173 (99.4)	115 (98.3)	.35	55 (96)	60 (100)	.14
Total number of household members, mean (SD)	3.6 (1.0)	3.7 (0.9)	.66	3.8 (1.0)	3.6 (0.7)	.17
Born in Norway, n (%)	154 (88.5)	96 (82.1)	.16	42 (75)	54 (90)	.03
Education: ≤4 years of college or university or lower, n (%)	67 (38.5)	65 (55.6)	.004	30 (53)	35 (58)	.54
Children						
Age (months), mean (SD)	10.8 (1.2)	11.0 (1.2)	.14	11.0 (1.4)	11.1 (1.1)	.60
Child's sex: female, n (%)	73 (42.0)	56 (47.9)	.32	27 (47)	29 (48)	.92
Children's food intake at baseline, mean (SD)						
Vegetables (13 items)						
Frequency (times/day)	3.2 (1.6)	3.0 (1.5)	.27	3.2 (1.4)	2.9 (1.6)	.35
Variation (number/week)	7.3 (2.6)	7.1 (2.7)	.47	7.3 (2.4)	6.9 (2.9)	.35
Fruits (11 items)						
Frequency (times/day)	2.9 (1.6)	2.8 (1.7)	.65	2.9 (1.4)	2.6 (1.9)	.36
Variation (number/week)	5.9 (2.2)	5.7 (2.3)	.47	5.8 (2.1)	5.6 (2.4)	.54
Discretionary foods (7 items)[f]						
Frequency (times/day)	0.7 (1.3)	0.9 (1.5)	.18	0.9 (1.4)	1.0 (1.6)	.87

[a]Participants who were lost to follow-up or had incomplete outcome data at follow-up 2.

[b]Calculated by the Pearson chi-square test or *t* test.

[c]Included co-mothers and foster mothers.

[d]N/A: not applicable; it was not relevant to the study to calculate *P* values for gender items.

[e]Lived together with the other parent.

[f]Included five unhealthy snack items and two sugar-sweetened beverages.

At the follow-up 1 time point, 71 participants were lost to follow-up and 11 had incomplete data on outcome variables. At the follow-up 2 time point, 109 were lost to follow-up and 8 had incomplete outcome data. The number of participants that completed the baseline questionnaire was 298, and 291 (97.7%) were included in our analyses. Of these, 209 (71.8%) completed the follow-up 1 questionnaire and 174 (59.8%) completed the follow-up 2 questionnaire.

Mean parental age at baseline was 31.7 years (SD 4.2) (see Table 1). Most participants were mothers (287/291, 98.6%), lived in two-parent households (288/291, 99.0%), and were born in Norway (250/290, 86.2%). Other characteristics that are not listed in the table are as follows: the mean age of the child at follow-up 1 was 17.8 months (SD 1.23; n=209) and at follow-up 2 was 24.2 months (SD 1.68; n=174). All 19 Norwegian counties were represented in the study sample. We observed a higher proportion of participants from the south of Norway compared to the national population data [53].

The infants had a frequency daily intake of 3.2 (SD 1.6) items of vegetables and 2.8 (SD 1.6) items of fruit. For discretionary

food, the weekly intake was less than 1 item (mean 0.8, SD 1.4) at baseline. The participating children ate a more varied range of vegetables (mean 7.2 per week, SD 2.6) compared to fruits (mean 5.8 per week, SD 2.2).

To get an overview of the baseline characteristics of participants who remained in the study and those who were lost to follow-up, Tables 2 and 3 present the baseline characteristics of these participants at the two follow-up time points. A comparison between all participants is presented, including how many of those lost to follow-up adhered to the intervention or control group. Of the 82 participants who did not participate in the follow-up 1 time point (see Table 2), 51 (62%) were from the intervention group and 31 (38%) were from the control group.

At follow-up 2, the number of nonresponders was comparable in the two groups (ie, 57/117, 48.7%, in the control group and 60/117, 51.3%, in the intervention group). Participants with a higher education level were more likely to complete the follow-up 2 questionnaires (P=.004).

Table 4 shows how many participants out of 144 in the intervention group completed each of the 22 lessons on the Food4toddlers website. The first two lessons were available when the participants got access to the website. After that, a new lesson was delivered every week. Lesson 1 was an information lesson (eg, how to navigate the website and information about the study), and lesson 7 had a gaming element included that was only accessible from a computer and not on mobile devices. Few parents completed this lesson (21/144, 14.6%). The number of parents out of 144 who completed lessons ranged from 21 (14.6%) to 87 (60.4%). We saw a general drop in parents completing lessons over time.

Table 4. Number of participants in the intervention group who completed lessons.

Lesson No.	Number of intervention group participants who completed each lesson (n=144), n (%)
1	68 (47.2)
2	87 (60.4)
3	78 (54.2)
4	70 (48.6)
5	63 (43.8)
6	60 (41.7)
7	21 (14.6)
8	60 (41.7)
9	63 (43.8)
10	53 (36.8)
11	46 (31.9)
12	52 (36.1)
13	54 (37.5)
14	59 (41.0)
15	52 (36.1)
16	49 (34.0)
17	45 (31.3)
18	45 (31.3)
19	45 (31.3)
20	36 (25.0)
21	32 (22.2)
22	29 (20.1)

Dietary Outcomes

At follow-up 1, a significant time × group interaction was observed for frequency of vegetable intake (P=.02); see adjusted measures in Table 5. The between-group difference in the change from baseline to follow-up 1 was 0.46 items per day (95% CI 0.06-0.86), showing a larger increase in the frequency of vegetable intake in the intervention group compared to the control group. No other significant differences in dietary changes from baseline to follow-up 1 or from baseline to follow-up 2 were observed between the groups.

Table 5. Intervention effects of the Food4toddlers study on food intake outcomes from baseline to both follow-up time points (n=291).

Intervention effects	Baseline to follow-up 1[a]		Baseline to follow-up 2[a]	
	Mean change estimate (95% CI)	*P* value	Mean change estimate (95% CI)	*P* value
Unadjusted				
Vegetables				
Frequency (times/day)	0.44 (0.04 to 0.84)	.03	0.30 (–0.14 to 0.74)	.18
Variation (number/week)	0.56 (–0.08 to 1.19)	.09	0.69 (0.04 to 1.43)	.06
Fruits				
Frequency (times/day)	0.04 (–0.45 to 0.54)	.87	–0.07 (–0.62 to 0.48)	.81
Variation (number/week)	0.07 (–0.51 to 0.66)	.80	–0.17 (–0.82 to 0.49)	.62
Discretionary foods				
Frequency (times/day)	–0.10 (–1.20 to 1.00)	.85	0.05 (–1.02 to 1.11)	.93
Adjusted[b]				
Vegetables				
Frequency (times/day)	0.46 (0.06 to 0.86)	.02	0.32 (–0.12 to 0.75)	.15
Variation (number/week)	0.60 (–0.04 to 1.23)	.07	0.73 (–0.01 to 1.46)	.05
Fruits				
Frequency (times/day)	0.03 (–0.47 to 0.52)	.91	–0.10 (–0.64 to 0.44)	.71
Variation (number/week)	0.09 (–0.50 to 0.67)	.78	–0.18 (–0.84 to 0.48)	.60
Discretionary foods				
Frequency (times/day)	–0.07 (–1.17 to 1.02)	.89	0.08 (–0.98 to 1.14)	.89

[a]Mean change in frequency or variety of vegetables, fruits, or discretionary foods from baseline to the postinterventions (follow-up 1 or 2) between the control and intervention groups.

[b]Adjusted for child age and gender and parental BMI, education level, and age at baseline.

The change in frequency of vegetable intake from baseline to the follow-up time points for the intervention group and the control group are presented in Figure 2A.

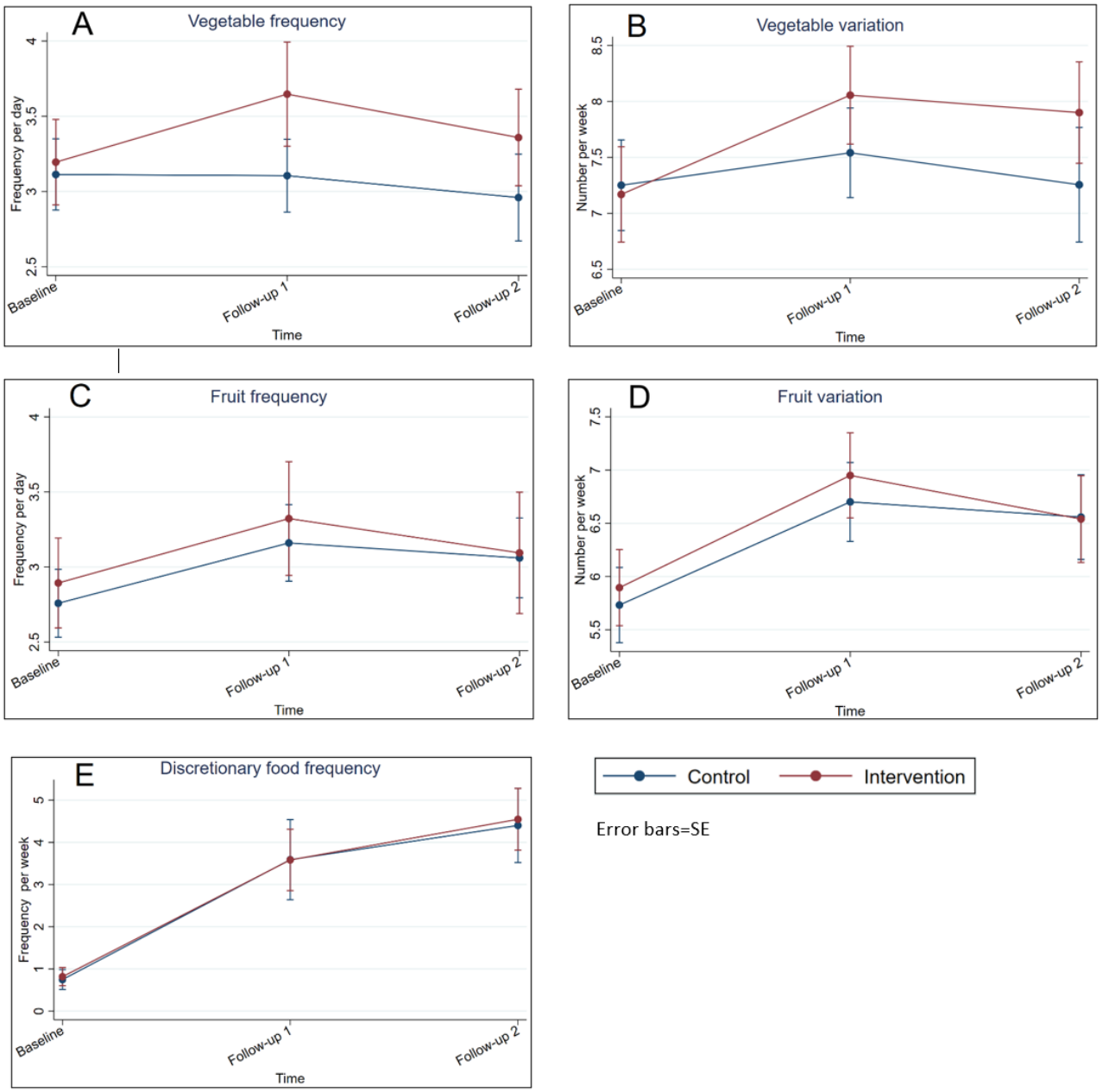
Estimated marginal means (EMMs) for the intervention group showed that the vegetable intake of 3.20 times per day (SE 0.15) at baseline increased to 3.65 times per day (SE 0.18) at follow-up 1. There was no change in frequency of vegetable intake in the control group from baseline (EMM 3.11 times per day, SE 0.12) to follow-up 1 (EMM 3.11 times per day, SE 0.12). A small decrease in frequency was observed from follow-up 1 to follow-up 2 in both groups: an EMM vegetable intake of 2.96 times per day (SE 0.15) was observed for the control group and 3.36 (SE 0.16) was observed for the intervention group. There was no significant time trend from baseline to the follow-up time points.

A similar trend was observed for the variety score of vegetables (see Figure 2B), although the group × time interactions were only borderline significant. Specifically, the group difference in the change from baseline to follow-up 1 was 0.60 vegetables tasted per week (*P*=.07) and from baseline to follow-up 2 was 0.73 vegetables tasted per week (*P*=.05) (see Table 5). Moreover, regarding vegetable variety, the EMM of the control group was 7.25 (SE 0.21) at baseline, 7.54 (SE 0.20) at follow-up 1, and 7.26 (SE 0.26) at follow-up 2; for the intervention group, the EMM was 7.17 (SE 0.22) at baseline, 8.06 (SE 0.22) at follow-up 1, and 7.90 (SE 0.23) at follow-up 2. No significant time trend from baseline to the follow-up time points was observed.

There were no significant between-group differences in change in the frequency nor variety of fruit intake from baseline to follow-up 1 and from baseline follow-up 2 (see Figure 2C and D). There was a significant time trend for fruit frequency from baseline to follow-up 1 (*P*=.002) and borderline significance from baseline to follow-up 2 (*P*=.052). For variety of fruit, a significant time trend was seen from baseline to both follow-up time points (*P*<.001).

No intervention effect was observed for the intake of discretionary foods, as there was no significant between-group difference in the change in intake frequency from baseline to either of the two follow-up time points. However, the intake of discretionary foods increased significantly (*P*<.001) over time (see Figure 2E). Specifically, at baseline, the intake was less than 1 item per week for both groups, which increased to 3.6 items per week in both groups (control EMM 3.6 items per week, SE 0.48; and intervention EMM 3.6 items per week, SE 0.37) at follow-up 1; this later increased to 4.4 (SE 0.45) items per week in the control group and 4.5 (SE 0.37) items per week in the intervention group at follow-up 2.

Figure 2. Estimated marginal means for children's food intake at baseline, follow-up 1, and follow-up 2 for the frequency of vegetables, fruits, and discretionary foods, and variety of vegetables and fruits. Values are adjusted for child age and gender and parental BMI, education level, and age reported at baseline.



The sensitivity analyses—*t* tests and Mann-Whitney *U* tests—using complete cases showed results in line with the GEE analyses, except for the results for vegetable variation, which were no longer borderline significant.

Discussion

Principal Findings

In this study, we observed that giving parents access to an eHealth intervention during toddlerhood increased their children's vegetable consumption frequency. The intervention effect was attenuated and no longer significant 6 months postintervention. A borderline significant effect for variety of vegetable intake in favor of the intervention group was observed at both time points. For fruits and discretionary foods, there were no intervention effects.

Although the intervention promoted a higher consumption of both vegetables and fruits, and lower consumption of discretionary foods, the vegetable promotion was the main focus in the Food4toddlers study, which may explain our findings. Specifically, the intervention focused on vegetables from the start; on the front page of the website and in the first lessons that the participants were able to access, vegetable promotion was central. Interest in the intervention website was highest at the start of the program, which was also seen in other web-based programs [54]. The video on the front page of the Food4toddlers website focused on how important just a small weekly increase in vegetable consumption may be for children's long-term health. It is possible that those who accessed the website

watched this video and that this may have motivated improved vegetable consumption. There may also be more room for improvement in vegetable intake relative to fruit or discretionary food intake in this age group. A British study found a larger variety of vegetables in commercially prepared dinners than in their home-cooked recipe counterparts [55]. Parents tend to serve more commercially prepared dinners at the age of 1 year than they do later [39,56]. This may explain why the variety of vegetables did not increase by age, even though children eat larger portions of foods by age [39,56]. A known reason for lack of variety is the age-specific trait of rejection of new foods—food neophobia—that peaks around 2 years of age [57]. In order to create a healthy eating environment, this study focused on the importance of repeated exposure as a means to enhance the acceptance of new foods before that age. There was a borderline significant difference between the groups for vegetable variety in our study, at both time points, that may indicate that children in the intervention group obtained a higher acceptance of these foods before the age of 2 years, which may persist over time [7]. While there are no studies directly comparable to this study, some have reported dietary outcomes of eHealth interventions targeting older or younger children. The Swedish *Mobile-based intervention intended to stop obesity in preschoolers* (MINISTOP) mobile health (mHealth) intervention [58] reported no effect on vegetable consumption. The MINISTOP study targeted parents of 4-year-olds, and the intervention group got access to an app for 6 months that focused on a healthy diet and physical activity. The results from the Australian Time2bHealthy study for vegetable consumption were in line with the MINISTOP study. Time2bHealthy delivered an 11-week web course on healthy lifestyle to the intervention group, followed by fortnightly emails for 3 months. They targeted parents of 2- to 5-year-old children with BMI values at or above the 50th percentile for their age, and the participants got individual feedback from a dietitian. Our intervention targeted parents with younger children in a period where dietary habits are established, which may explain the positive results for vegetables in our study. A Norwegian eHealth RCT intervention, *Early food for future health*, delivered monthly videos on child-feeding to parents of infants (6-12 months of age) [59] and found an intervention effect for vegetable variation [59]. They made a composite score of fruit and vegetable frequency, which also showed improvement in intake [59]. A similar score was used in two studies targeting older children [60,61] showing positive results, contrary to no reported effect found in an mHealth study targeting infants [20]. A composite healthy lifestyle score was assessed in the MINISTOP study that showed a positive intervention effect [58].

Both the intervention and the control groups increased their intake of fruit over time, but no intervention effects between the groups were seen. The lack of effect on fruit intake contrary to vegetable intake may be explained by the differences between the two types of foods in terms of skills and time needed for preparation, consumption patterns, and the parents' readiness to make changes [62-64]. Few preparations are necessary to give the child a fruit as a snack or in a smoothie, and fruits are more easily accepted by children than vegetables due to their sweet taste [65]. The children may have tasted and accepted a variety of fruits before the intervention period started, and improvements may be hard to obtain. The lack of an intervention effect on fruit consumption has been observed in comparable studies [58,59,66].

In contrast with our findings and those from other studies [59,67], the Time2bHealthy study showed an effect on discretionary foods in favor of the intervention group [66]. A review exploring both traditional and eHealth interventions aiming to reduce sugar - sweetened beverages among young children (<5 years of age) found that success was more likely if interventions were multicomponent, targeted vulnerable populations, and had a high intervention intensity and contact time [68]. The Time2bHealthy study was conducted in line with these success factors, which may explain the positive results. The MINISTOP study found an intervention effect on sweetened beverage consumption in favor of the intervention group [67]. The offering of discretionary foods was low at baseline in the Food4toddlers study and increased at the follow-up time points in both groups. However, the intake remained relatively low when compared with other studies [69,70]. The increase in both groups over time may be explained by the fact that children tend to incorporate the rest of their family's eating patterns, including more discretionary food, during the second year of life (eg, ice cream in the summer and biscuits as snacks).

Findings from this study and other similar studies show that digital interventions may be effective in improving some aspects of dietary intake. However, for most parent-focused eHealth studies, long-term retention of effects have not been observed [17,58]. One interesting exception is the long-term effect on discretionary foods in the Time2bHealthy study [66]. The lack of long-term effects is a challenge for eHealth interventions aimed at lifestyle behavior, in general [71], and specifically for parent-focused, traditional and online, obesity prevention interventions [13,72]. A duration of 6 months or shorter is common in parent-focused eHealth interventions [13,25,73]. A longer duration might contribute to maintained effects over time [13,71,74]. Further, including short and thematically narrow "booster sessions" after the end of more intensive intervention sessions have shown promising results [72] and may also maintain the effects of the intervention. Such short booster sessions have a low participant burden, can be important reminders, and can easily be conducted in eHealth interventions. A review showed that combining web-delivered interventions with other delivery modes, such as SMS, telephone coaching, and emails, had stronger effects on behavior changes over time [21]. The process evaluation of this study [75] showed that 13% of the invited participants did not enter the Food4toddlers website at all, indicating some challenges in engaging all participants. Other deliveries might have been valuable toward achieving better engagement. However, personal contact is cost- and time-consuming, which limits distribution to the population at large [71]. Digital tailoring based on information about diet and physical activity provided by parents on the website or app, as done in the MINISTOP study [76], may contribute to better adherence. Even though the effect did not last after follow-up 1 in this study, there is still a possible public health benefit of

increasing vegetable intake among children, even in small measures.

Strengths and Limitations

Few parent-focused eHealth studies are exclusively web based and few target young children [13]. The participants in the Food4toddlers study represented all 19 Norwegian counties, which was possible because we used Facebook as the recruitment platform and had no face-to-face components in the intervention. The possibility of reaching a large and widespread population is one of the main benefits of using eHealth approaches [21]; however, we aimed for a larger sample in this study. Separate analyses for fruits and vegetables could also be viewed as a strength due to different consumption patterns and tastes [63-65] and are recommended for studies targeting young children [77,78]. A recently published review paper addressed the need to examine both variety and intake (ie, quantity) of fruits and vegetables due to the different findings regarding health outcomes; this review also revealed that such research was particularly lacking in young age groups [35].

A limitation of the study is the low generalizability of the findings due to the participants' education level, which was higher than national figures [79]. It is conceivable that a more representative sample might have resulted in a larger intervention effect, as indicated in other studies [80,81]. Even though both parents were invited to participate, 287 of 291 (98.6%) participants were mothers. We do not know if our findings would have been different if more fathers were included. We aimed at recruiting a larger sample, but time and cost (ie, expensive Facebook advertisement) limited that. Therefore, we ended up with a more restricted sample size and, hence, lower statistical power than was planned for. It turned out to be challenging to recruit parents through Facebook when the children were around 10 months of age, possibly because parents in Norway often start working after their maternity leave around that time. Quantifying the dietary intake in grams and nutrient calculation might have added value to the assessments; however, portion size estimations were not recorded. Self-reported FFQs have limitations, especially in this age group where dietary habits are rapidly changing and the answers are solely dependent on the parents' observations and suggestions [82]. A potential bias in intervention group reporting could be answering according to the perceived intention of the intervention (eg, higher intake of vegetables) [83]. The three questionnaires were delivered in different Nordic seasons—two in autumn-winter and one in winter-spring—which could have influenced the results, especially for fruit and vegetable intake. If so, the effect of the intervention would tend to be overrated. The digital approach limited the possibility of collecting objective measurements, leaving self-reported measures as the only option, which have limitations [84].

Conclusions

In this study, we investigated the effects of the Norwegian Food4toddlers RCT. An intervention effect on the frequency of intake of vegetables was observed immediately after the 6-month intervention period ended. The difference was attenuated and no longer significant at the follow-up 2 time point, 6 months postintervention. The consumption of discretionary food increased over time in both groups.

Despite the potential of reaching a large population with limited resources, few eHealth interventions seeking to enhance children's diets have targeted parents of toddlers at this key time in children's food preference development. Our results show that there is a potential to improve aspects of young children's diets utilizing this kind of intervention. To obtain long-term effects in eHealth interventions, longer durations should be considered along with tailoring in a digital or a personal form. Delivering short reminders after the end of the main content of the intervention may contribute to better adherence and is feasible in eHealth interventions.

Acknowledgments

The authors wish to thank the participants and the students who were involved in the development of the intervention, in particular, Svein Even Skogen, Christina Lien, Ida Kile Næss, Anne Kate Edvardsen, and Kaia Heslien. This study was funded by the University of Agder, Norway. The financial contributor was not involved in designing the study; in the collection, analyses, and interpretation of data; or in writing the manuscript.

Authors' Contributions

MR, EH, FV, and NØ initiated and designed the study and developed the intervention. MR performed the data collection with supervision by EH, FV, and NØ. MR and AM performed the analyses. MR drafted the manuscript with substantial input from AM, EH, FV, WVL, KC, and NØ. All authors contributed to, read, and approved the final version of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH V1.6.1 checklist for the Food4toddlers study.
[PDF File (Adobe PDF File), 1543 KB - jmir_v23i2e18311_app1.pdf]

References

1. Johnson BJ, Hendrie GA, Golley RK. Reducing discretionary food and beverage intake in early childhood: A systematic review within an ecological framework. Public Health Nutr 2015 Oct 21;19(9):1684-1695. [doi: 10.1017/s1368980015002992] [Medline: 26487544]
2. McCarthy R, Kehoe L, Flynn A, Walton J. The role of fruit and vegetables in the diets of children in Europe: Current state of knowledge on dietary recommendations, intakes and contribution to energy and nutrient intakes. In: Proceedings of the 13th European Nutrition Conference. 2020 Aug 13 Presented at: The 13th European Nutrition Conference; October 15-18, 2019; Dublin, Ireland p. 479-486. [doi: 10.1017/s0029665120007090]
3. Hodder R, O'Brien K, Tzelepis F, Wyse R, Wolfenden L. Interventions for increasing fruit and vegetable consumption in children aged five years and under. Cochrane Database Syst Rev 2020 May 25;2020(5):1-407. [doi: 10.1002/14651858.cd008552.pub7] [Medline: 32449203]
4. WHO technical staff. Increasing fruit and vegetable consumption to reduce the risk of noncommunicable diseases. World Health Organization. 2014 Sep. URL: https://www.who.int/elena/titles/bbc/fruit_vegetables_ncds/en/ [accessed 2021-02-04]
5. Report of the Commission on Ending Childhood Obesity. Geneva, Switzerland: World Health Organization; 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/204176/9789241510066_eng.pdf?sequence=1 [accessed 2021-02-04]
6. Aune D, Giovannucci E, Boffetta P, Fadnes L, Keum N, Norat T. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality-A systematic review and dose-response meta-analysis of prospective studies. Int J Epidemiol 2017;46(3):1029-1056. [doi: 10.1093/ije/dyw319] [Medline: 28338764]
7. Scaglioni S, De Cosmi V, Ciappolino V, Parazzini F, Brambilla P, Agostoni C. Factors influencing children's eating behaviours. Nutrients 2018 May 31;10(6):706. [doi: 10.3390/nu10060706] [Medline: 29857549]
8. Grimm KA, Kim SA, Yaroch AL, Scanlon KS. Fruit and vegetable intake during infancy and early childhood. Pediatrics 2014 Sep;134 Suppl 1:S63-S69 [FREE Full text] [doi: 10.1542/peds.2014-0646K] [Medline: 25183758]
9. Rose CM, Birch LL, Savage JS. Dietary patterns in infancy are associated with child diet and weight outcomes at 6 years. Int J Obes 2017 Feb 28;41(5):783-788. [doi: 10.1038/ijo.2017.27] [Medline: 28133360]
10. Bjelland M, Brantsæter AL, Haugen M, Meltzer HM, Nystad W, Andersen LF. Changes and tracking of fruit, vegetables and sugar-sweetened beverages intake from 18 months to 7 years in the Norwegian mother and child cohort study. BMC Public Health 2013 Aug 30;13(1):1-11. [doi: 10.1186/1471-2458-13-793] [Medline: 24103398]
11. Larsen JK, Hermans RC, Sleddens EF, Engels RC, Fisher JO, Kremers SP. How parental dietary behavior and food parenting practices affect children's dietary behavior. Interacting sources of influence? Appetite 2015 Jun;89:246-257. [doi: 10.1016/j.appet.2015.02.012] [Medline: 25681294]
12. Faith MS, Van Horn L, Appel LJ, Burke LE, Carson JAS, Franch HA, et al. Evaluating parents and adult caregivers as "agents of change" for treating obese children: Evidence for parent behavior change strategies and research gaps. Circulation 2012 Mar 06;125(9):1186-1207. [doi: 10.1161/cir.0b013e31824607ee] [Medline: 22271754]
13. Hammersley ML, Jones RA, Okely AD. Parent-focused childhood and adolescent overweight and obesity eHealth interventions: A systematic review and meta-analysis. J Med Internet Res 2016 Jul 21;18(7):e203 [FREE Full text] [doi: 10.2196/jmir.5893] [Medline: 27443862]
14. Redsell SA, Edmonds B, Swift JA, Siriwardena AN, Weng S, Nathan D, et al. Systematic review of randomised controlled trials of interventions that aim to reduce the risk, either directly or indirectly, of overweight and obesity in infancy and early childhood. Matern Child Nutr 2015 Apr 20;12(1):24-38. [doi: 10.1111/mcn.12184] [Medline: 25894857]
15. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: A systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]
16. Slomian J, Bruyère O, Reginster J, Emonts P. The internet as a source of information used by women after childbirth to meet their need for information: A web-based survey. Midwifery 2017 May;48:46-52. [doi: 10.1016/j.midw.2017.03.005] [Medline: 28324809]
17. Helle C, Hillesund ER, Wills AK, Øverby NC. Examining the effects of an eHealth intervention from infant age 6 to 12 months on child eating behaviors and maternal feeding practices one year after cessation: The Norwegian randomized controlled trial Early Food for Future Health. PLoS ONE 2019 Aug 23;14(8):e0220437. [doi: 10.1371/journal.pone.0220437] [Medline: 31442241]
18. Moon RY, Mathews A, Oden R, Carlin R. Mothers' perceptions of the internet and social media as sources of parenting and health information: Qualitative study. J Med Internet Res 2019 Jul 09;21(7):e14289 [FREE Full text] [doi: 10.2196/14289] [Medline: 31290403]
19. Litterbach E, Russell CG, Taki S, Denney-Wilson E, Campbell KJ, Laws RA. Factors influencing engagement and behavioral determinants of infant feeding in an mHealth program: Qualitative evaluation of the growing healthy program. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e196 [FREE Full text] [doi: 10.2196/mhealth.8515] [Medline: 29254908]
20. Russell CG, Denney-Wilson E, Laws RA, Abbott G, Zheng M, Lymer SJ, et al. Impact of the growing healthy mHealth program on maternal feeding practices, infant food preferences, and satiety responsiveness: Quasi-experimental study. JMIR Mhealth Uhealth 2018 Apr 25;6(4):e77 [FREE Full text] [doi: 10.2196/mhealth.9303] [Medline: 29695373]
21. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]

22. Birch LL, Ventura AK. Preventing childhood obesity: What works? Int J Obes 2009 Apr 13;33(S1):S74-S81. [doi: 10.1038/ijo.2009.22] [Medline: 19363514]
23. Pearson N, Biddle SJ, Gorely T. Family correlates of fruit and vegetable consumption in children and adolescents: A systematic review. Public Health Nutr 2009 Feb 01;12(2):267-283. [doi: 10.1017/s1368980008002589] [Medline: 18559129]
24. Lacy KE, Spence AC, McNaughton SA, Crawford DA, Wyse RJ, Wolfenden L, et al. Home environment predictors of vegetable and fruit intakes among Australian children aged 18 months. Appetite 2019 Aug;139:95-104. [doi: 10.1016/j.appet.2019.04.009] [Medline: 30991083]
25. Røed M, Hillesund ER, Vik FN, Van Lippevelde W, Øverby NC. The Food4toddlers study - Study protocol for a web-based intervention to promote healthy diets for toddlers: A randomized controlled trial. BMC Public Health 2019 May 14;19(1):563. [doi: 10.1186/s12889-019-6915-x] [Medline: 31088438]
26. Mat til minsten (Food4toddlers). University of Agder. 2021. URL: https://www.uia.no/mattilminsten [accessed 2021-02-04]
27. SurveyXact. Rambøll. 2019. URL: https://www.surveyxact.no/ [accessed 2021-02-04]
28. Brug J, Oenema A, Ferreira I. Theory, evidence and intervention mapping to improve behavior nutrition and physical activity interventions. Int J Behav Nutr Phys Act 2005 Apr 04;2(1):2 [FREE Full text] [doi: 10.1186/1479-5868-2-2] [Medline: 15807898]
29. Kelder S, Hoelscher D, Perry C. How individuals, environments, and health behaviors interact. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior: Theory, Research, and Practice. 5th edition. San Francisco, CA: Jossey-Bass; 2015:159-181.
30. Petty R, Barden J, Wheeler S. The elaboration likelihood model of persuasion: Developing health promotions for sustained behavioral change. In: DiClemente RJ, Crosby RA, Kegler MC, editors. Emerging Theories in Health Promotion Practice and Research. 2nd edition. San Francisco, CA: Jossey-Bass; 2009:185-214.
31. Mak TN, Prynne CJ, Cole D, Fitt E, Bates B, Stephen AM. Patterns of sociodemographic and food practice characteristics in relation to fruit and vegetable consumption in children: Results from the UK National Diet and Nutrition Survey Rolling Programme (2008–2010). Public Health Nutr 2013 Aug 07;16(11):1912-1923. [doi: 10.1017/s1368980013001912] [Medline: 23920151]
32. Jacobs JD, Steffen L. Nutrients, foods, and dietary patterns as exposures in research: A framework for food synergy. Am J Clin Nutr 2003 Sep;78(3 Suppl):508S-513S. [doi: 10.1093/ajcn/78.3.508S] [Medline: 12936941]
33. Cena H, Calder PC. Defining a healthy diet: Evidence for the role of contemporary dietary patterns in health and disease. Nutrients 2020 Jan 27;12(2):334. [doi: 10.3390/nu12020334] [Medline: 32012681]
34. Wirt A, Collins CE. Diet quality – What is it and does it matter? Public Health Nutr 2009 Apr 01;12(12):2473-2492. [doi: 10.1017/s136898000900531x] [Medline: 19335941]
35. Marshall AN, van den Berg A, Ranjit N, Hoelscher DM. A scoping review of the operationalization of fruit and vegetable variety. Nutrients 2020 Sep 19;12(9):2868. [doi: 10.3390/nu12092868] [Medline: 32961721]
36. Ramsay SA, Shriver LH, Taylor CA. Variety of fruit and vegetables is related to preschoolers' overall diet quality. Prev Med Rep 2017 Mar;5:112-117 [FREE Full text] [doi: 10.1016/j.pmedr.2016.12.003] [Medline: 27990339]
37. Røed M, Vik FN, Hillesund ER, Lippevelde WV, Øverby NC. Associations between parental food choice motives, health-promoting feeding practices, and infants' fruit and vegetable intakes: The Food4toddlers study. Food Nutr Res 2020;64:3730 [FREE Full text] [doi: 10.29219/fnr.v64.3730] [Medline: 33240032]
38. Norwegian Mother, Father and Child Cohort Study (MoBa). Norwegian Institute of Public Health. 2018. URL: https://www.fhi.no/en/studies/moba/ [accessed 2021-02-04]
39. Øverby N, Kristiansen A, Andersen L, Lande B. Spedkost 12 Months: National Dietary Survey Among 12-Month-Old Children. Oslo, Norway: Norwegian Directorate of Health; 2009 Mar. URL: https://tinyurl.com/y0sofp8l [accessed 2021-02-04]
40. Andersen L, Lande B, Trygg K, Hay G. Validation of a semi-quantitative food-frequency questionnaire used among 2-year-old Norwegian children. Public Health Nutr 2007 Jan 02;7(6):757-764. [doi: 10.1079/phn2004613] [Medline: 15369614]
41. Jin F. Questionnaire 5 When the Child Was 18 Months Old. The Norwegian Mother and Child Cohort Study (MoBa): Mother Questionnaire. Version 1.0. Oslo, Norway: Norwegian Institute of Public Health; 2016 Mar. URL: https://www.fhi.no/globalassets/dokumenterfiler/studier/moba/dokumenter/instrument-documentation-q5.pdf [accessed 2021-02-04]
42. Survey of consumer expenditure. Statistics Norway. 2019. URL: https://www.ssb.no/en/statbank/table/10249 [accessed 2021-02-04]
43. Campbell KJ, Crawford DA, Ball K. Family food environment and dietary behaviors likely to promote fatness in 5-6-year-old children. Int J Obes 2006 Feb 21;30(8):1272-1280. [doi: 10.1038/sj.ijo.0803266] [Medline: 16491108]
44. Kiefner-Burmeister AE, Hoffmann DA, Meers MR, Koball AM, Musher-Eizenman DR. Food consumption by young children: A function of parental feeding goals and practices. Appetite 2014 Mar;74:6-11. [doi: 10.1016/j.appet.2013.11.011] [Medline: 24275668]
45. Harrison K, Liechty JM. US preschoolers' media exposure and dietary habits: The primacy of television and the limits of parental mediation. J Child Media 2012 Feb;6(1):18-36. [doi: 10.1080/17482798.2011.633402]

46. Lanfer A, Hebestreit A, Ahrens W, Krogh V, Sieri S, Lissner L, IDEFICS Consortium. Reproducibility of food consumption frequencies derived from the Children's Eating Habits Questionnaire used in the IDEFICS study. Int J Obes (Lond) 2011 Apr;35 Suppl 1:S61-S68. [doi: 10.1038/ijo.2011.36] [Medline: 21483424]
47. Cutler DM, Lleras-Muney A. Education and Health: Evaluating Theories and Evidence. Cambridge, MA: National Bureau of Economic Research; 2006 Jun. URL: https://www.nber.org/system/files/working_papers/w12352/w12352.pdf [accessed 2021-02-05]
48. Angelopoulos P, Kourlaba G, Kondaki K, Fragiadakis GA, Manios Y. Assessing children's diet quality in Crete based on Healthy Eating Index: The Children Study. Eur J Clin Nutr 2009 Aug;63(8):964-969. [doi: 10.1038/ejcn.2009.10] [Medline: 19223917]
49. Salazar A, Ojeda B, Dueñas M, Fernández F, Failde I. Simple generalized estimating equations (GEEs) and weighted generalized estimating equations (WGEEs) in longitudinal studies with dropouts: Guidelines and implementation in R. Stat Med 2016 Aug 30;35(19):3424-3448. [doi: 10.1002/sim.6947] [Medline: 27059703]
50. Rasmussen M, Krølner R, Klepp KI, Lytle L, Brug J, Bere E, et al. Determinants of fruit and vegetable consumption among children and adolescents: A review of the literature. Part I: Quantitative studies. Int J Behav Nutr Phys Act 2006 Aug 11;3:22 [FREE Full text] [doi: 10.1186/1479-5868-3-22] [Medline: 16904006]
51. Bell ML, Fiero M, Horton NJ, Hsu C. Handling missing data in RCTs: A review of the top medical journals. BMC Med Res Methodol 2014 Nov 19;14:118 [FREE Full text] [doi: 10.1186/1471-2288-14-118] [Medline: 25407057]
52. University of Agder. UiA Open Research Data. University of Agder. 2020. URL: https://dataverse.no/dataverse/uia [accessed 2021-02-01]
53. Population. Statistics Norway. 2019. URL: https://www.ssb.no/en/befolkning/statistikker/folkemengde [accessed 2021-02-04]
54. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
55. Carstairs SA, Craig LC, Marais D, Bora OE, Kiezebrink K. A comparison of preprepared commercial infant feeding meals with home-cooked recipes. Arch Dis Child 2016 Jul 19;101(11):1037-1042. [doi: 10.1136/archdischild-2015-310098] [Medline: 27435891]
56. Kristiansen A, Andersen L, Lande B. Småbarnskost 2 år: National Dietary Survey Among 2-Year-Old Children. Oslo, Norway: Norwegian Directorate of Health; 2009 Aug. URL: https://tinyurl.com/7vk8mfux [accessed 2021-02-04]
57. Dovey TM, Staples PA, Gibson EL, Halford JC. Food neophobia and 'picky/fussy' eating in children: A review. Appetite 2008 Mar;50(2-3):181-193. [doi: 10.1016/j.appet.2007.09.009] [Medline: 17997196]
58. Delisle Nyström C, Sandin S, Henriksson P, Henriksson H, Maddison R, Löf M. A 12-month follow-up of a mobile-based (mHealth) obesity prevention intervention in pre-school children: The MINISTOP randomized controlled trial. BMC Public Health 2018 May 24;18(1):658. [doi: 10.1186/s12889-018-5569-4] [Medline: 29793467]
59. Helle C, Hillesund ER, Wills AK, Øverby NC. Evaluation of an eHealth intervention aiming to promote healthy food habits from infancy-The Norwegian randomized controlled trial Early Food for Future Health. Int J Behav Nutr Phys Act 2019 Jan 3;16(1):1. [doi: 10.1186/s12966-018-0763-4] [Medline: 30606197]
60. Knowlden AP, Conrad E. Two-year outcomes of the Enabling Mothers to Prevent Pediatric Obesity Through Web-Based Education and Reciprocal Determinism (EMPOWER) randomized control trial. Health Educ Behav 2017 Sep 27;45(2):262-276. [doi: 10.1177/10901981177326O4] [Medline: 28954544]
61. Chen J, Weiss S, Heyman MB, Cooper B, Lustig RH. The efficacy of the web-based childhood obesity prevention program in Chinese American adolescents (Web ABC study). J Adolesc Health 2011 Aug;49(2):148-154 [FREE Full text] [doi: 10.1016/j.jadohealth.2010.11.243] [Medline: 21783046]
62. Appleton KM, Hemingway A, Saulais L, Dinnella C, Monteleone E, Depezay L, et al. Increasing vegetable intakes: Rationale and systematic review of published interventions. Eur J Nutr 2016 Jan 11;55(3):869-896. [doi: 10.1007/s00394-015-1130-8] [Medline: 26754302]
63. Glasson C, Chapman K, James E. Fruit and vegetables should be targeted separately in health promotion programmes: Differences in consumption levels, barriers, knowledge and stages of readiness for change. Public Health Nutr 2010 Jun 25;14(4):694-701. [doi: 10.1017/s1368980010001643] [Medline: 20576194]
64. Appleton K, Hemingway A, Rajska J, Hartwell H. Repeated exposure and conditioning strategies for increasing vegetable liking and intake: Systematic review and meta-analyses of the published literature. Am J Clin Nutr 2018 Oct 01;108(4):842-856 [FREE Full text] [doi: 10.1093/ajcn/nqy143] [Medline: 30321277]
65. Beauchamp GK, Mennella JA. Early flavor learning and its impact on later feeding behavior. J Pediatr Gastroenterol Nutr 2009 Mar;48 Suppl 1:S25-S30. [doi: 10.1097/MPG.0b013e31819774a5] [Medline: 19214055]
66. Hammersley ML, Okely AD, Batterham MJ, Jones RA. An internet-based childhood obesity prevention program (Time2bHealthy) for parents of preschool-aged children: Randomized controlled trial. J Med Internet Res 2019 Feb 08;21(2):e11964 [FREE Full text] [doi: 10.2196/11964] [Medline: 30735139]
67. Nyström CD, Sandin S, Henriksson P, Henriksson H, Trolle-Lagerros Y, Larsson C, et al. Mobile-based intervention intended to stop obesity in preschool-aged children: The MINISTOP randomized controlled trial. Am J Clin Nutr 2017 Apr 26;105(6):1327-1335. [doi: 10.3945/ajcn.116.150995] [Medline: 28446496]

68. Vercammen KA, Frelier JM, Lowery CM, McGlone ME, Ebbeling CB, Bleich SN. A systematic review of strategies to reduce sugar-sweetened beverage consumption among 0-year to 5-year olds. Obes Rev 2018 Nov;19(11):1504-1524. [doi: 10.1111/obr.12741] [Medline: 30019442]
69. van Grieken A, Vlasblom E, Wang L, Beltman M, Boere-Boonekamp MM, L'Hoir MP, et al. Personalized web-based advice in combination with well-child visits to prevent overweight in young children: Cluster randomized controlled trial. J Med Internet Res 2017 Jul 27;19(7):e268 [FREE Full text] [doi: 10.2196/jmir.7115] [Medline: 28751299]
70. Webb KL, Lahti-Koski M, Rutishauser I, Hector DJ, Knezevic N, Gill T, et al. Consumption of 'extra' foods (energy-dense, nutrient-poor) among children aged 16–24 months from Western Sydney, Australia. Public Health Nutr 2006 Dec 01;9(8):1035-1044. [doi: 10.1017/phn2006970] [Medline: 17125568]
71. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: A systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
72. Yavuz HM, van Ijzendoorn MH, Mesman J, van der Veek S. Interventions aimed at reducing obesity in early childhood: A meta-analysis of programs that involve parents. J Child Psychol Psychiatr 2014 Oct 08;56(6):677-692. [doi: 10.1111/jcpp.12330] [Medline: 25292319]
73. Helle C, Hillesund ER, Omholt ML, Øverby NC. Early food for future health: A randomized controlled trial evaluating the effect of an eHealth intervention aiming to promote healthy food habits from early childhood. BMC Public Health 2017 Sep 20;17(1):729. [doi: 10.1186/s12889-017-4731-8] [Medline: 28931384]
74. Verjans-Janssen SRB, van de Kolk I, Van Kann DHH, Kremers SPJ, Gerards SMPL. Effectiveness of school-based physical activity and nutrition interventions with direct parental involvement on children's BMI and energy balance-related behaviors – A systematic review. PLoS ONE 2018 Sep 27;13(9):e0204560. [doi: 10.1371/journal.pone.0204560] [Medline: 30261057]
75. Røed M, Vik F, Hillesund E, Van Lippevelde W, Medin A, Øverby NC. Process evaluation of an eHealth intervention (Food4toddlers) to improve toddlers' diet: Randomized controlled trial. JMIR Hum Factors 2020 Jul 06;7(3):e18171 [FREE Full text] [doi: 10.2196/18171] [Medline: 32628612]
76. Delisle C, Sandin S, Forsum E, Henriksson H, Trolle-Lagerros Y, Larsson C, et al. A web- and mobile phone-based intervention to prevent obesity in 4-year-olds (MINISTOP): A population-based randomized controlled trial. BMC Public Health 2015 Feb 7;15(1):95. [doi: 10.1186/s12889-015-1444-8] [Medline: 25886009]
77. Krølner R, Rasmussen M, Brug J, Klepp K, Wind M, Due P. Determinants of fruit and vegetable consumption among children and adolescents: A review of the literature. Part II: Qualitative studies. Int J Behav Nutr Phys Act 2011 Oct 14;8:112 [FREE Full text] [doi: 10.1186/1479-5868-8-112] [Medline: 21999291]
78. Ramsay SA, Eskelsen AK, Branen LJ, Armstrong Shultz J, Plumb J. Nutrient intake and consumption of fruit and vegetables in young children. Child Obes Nutr 2014 Sep 22;6(6):332-344 [FREE Full text] [doi: 10.1177/1941406414549622]
79. Educational attainment of the population. Statistics Norway. 2018. URL: https://www.ssb.no/en/utniv/ [accessed 2021-02-04]
80. Pinket A, De Craemer M, Huybrechts I, De Bourdeaudhuij I, Deforche B, Cardon G, et al. Diet quality in European pre-schoolers: Evaluation based on diet quality indices and association with gender, socio-economic status and overweight, the ToyBox-study. Public Health Nutr 2016 Apr 18;19(13):2441-2450. [doi: 10.1017/s1368980016000604] [Medline: 27087125]
81. Spence AC, Campbell KJ, Lioret S, McNaughton SA. Early childhood vegetable, fruit, and discretionary food intakes do not meet dietary guidelines, but do show socioeconomic differences and tracking over time. J Acad Nutr Diet 2018 Sep;118(9):1634-1643.e1. [doi: 10.1016/j.jand.2017.12.009] [Medline: 29482964]
82. Lovell A, Bulloch R, Wall CR, Grant CC. Quality of food-frequency questionnaire validation studies in the dietary assessment of children aged 12 to 36 months: A systematic literature review. J Nutr Sci 2017 May 08;6:e16. [doi: 10.1017/jns.2017.12] [Medline: 28630693]
83. Delgado-Rodríguez M, Llorca J. Bias. J Epidemiol Community Health 2004 Aug;58(8):635-641 [FREE Full text] [doi: 10.1136/jech.2003.008466] [Medline: 15252064]
84. Schoeller DA. Limitations in the assessment of dietary energy intake by self-report. Metabolism 1995 Feb;44(2 Suppl 2):18-22. [doi: 10.1016/0026-0495(95)90204-x] [Medline: 7869932]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth
EMM: estimated marginal mean
FFQ: food frequency questionnaire
GEE: generalized estimating equation
ISRCTN: International Standard Randomised Controlled Trial Number
mHealth: mobile health
MINISTOP: Mobile-based intervention intended to stop obesity in preschoolers
RCT: randomized controlled trial
WHO: World Health Organization

Edited by G Eysenbach, R Kukafka; submitted 21.02.20; peer-reviewed by M Laureati, C Delisle; comments to author 29.08.20; revised version received 08.12.20; accepted 17.01.21; published 16.02.21.

Please cite as:
Røed M, Medin AC, Vik FN, Hillesund ER, Van Lippevelde W, Campbell K, Øverby NC
Effect of a Parent-Focused eHealth Intervention on Children's Fruit, Vegetable, and Discretionary Food Intake (Food4toddlers): Randomized Controlled Trial
J Med Internet Res 2021;23(2):e18311
URL: http://www.jmir.org/2021/2/e18311/
doi:10.2196/18311
PMID:33591279

Original Paper

Peer-to-Peer Sharing of Social Media Messages on Sexual Health in a School-Based Intervention: Opportunities and Challenges Identified in the STASH Feasibility Trial

Maija Hirvonen[1], MSc; Carrie Purcell[2], PhD; Lawrie Elliott[3], PhD; Julia V Bailey[4], MD; Sharon Anne Simpson[2], PhD; Lisa McDaid[2], PhD; Laurence Moore[2], PhD; Kirstin Rebecca Mitchell[2], PhD; The STASH Study Team[5]

[1]Department of Public Health and Policy, London School of Hygiene and Tropical Medicine (see Acknowledgments), London, United Kingdom
[2]MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Glasgow, United Kingdom
[3]Department of Nursing and Community Health, Glasgow Caledonian University, Glasgow, United Kingdom
[4]Department of Primary Care and Population Health, University College London, London, United Kingdom
[5]see Acknowlegments

Corresponding Author:
Kirstin Rebecca Mitchell, PhD
MRC/CSO Social and Public Health Sciences Unit
University of Glasgow
Berkeley Square
99 Berkeley St
Glasgow, G3 7HR
United Kingdom
Phone: 44 1413537500
Email: kirstin.mitchell@glasgow.ac.uk

Abstract

Background: There is a strong interest in the use of social media to spread positive sexual health messages through social networks of young people. However, research suggests that this potential may be limited by a reluctance to be visibly associated with sexual health content on the web or social media and by the lack of trust in the veracity of peer sources.

Objective: The aim of this study was to investigate opportunities and challenges of using social media to facilitate peer-to-peer sharing of sexual health messages within the context of STASH (Sexually Transmitted Infections and Sexual Health), a secondary school-based and peer-led sexual health intervention.

Methods: Following training, and as a part of their role, student-nominated peer supporters (aged 14-16 years) invited school friends to trainer-monitored, private Facebook groups. Peer supporters posted curated educational sex and relationship content within these groups. Data came from a feasibility study of the STASH intervention in 6 UK schools. To understand student experiences of the social media component, we used data from 11 semistructured paired and group interviews with peer supporters and their friends (collectively termed students; n=42, aged 14-16 years), a web-based postintervention questionnaire administered to peer supporters (n=88), and baseline and follow-up questionnaires administered to students in the intervention year group (n=680 and n=603, respectively). We carried out a thematic analysis of qualitative data and a descriptive analysis of quantitative data.

Results: Message sharing by peer supporters was hindered by variable engagement with Facebook. The trainer-monitored and private Facebook groups were acceptable to student members (peer supporters and their friends) and reassuring to peer supporters but led to engagement that ran parallel to—rather than embedded in—their routine social media use. The offline context of a school-based intervention helped legitimate and augment Facebook posts; however, even where friends were receptive to STASH messages, they did not necessarily engage visibly on social media. Preferences for content design varied; however, humor, color, and text brevity were important. Preferences for social media versus offline message sharing varied.

Conclusions: Invitation-only social media groups formed around peer supporters' existing friendship networks hold potential for diffusing messages in peer-based sexual health interventions. Ideally, interactive opportunities should not be limited to single social media platforms and should run alongside offline conversations. There are tensions between offering young people autonomy to engage flexibly and authentically and the need for adult oversight of activities for information accuracy and safeguarding.

(J Med Internet Res 2021;23(2):e20898) doi:10.2196/20898

KEYWORDS

social media; sexual health; sex education; peer education; process evaluation; school; feasibility studies; adolescent; social networking

Introduction

Young people (defined by the World Health Organization as 10-24 years) use social media as primary channels for digital interaction [1], and there is burgeoning interest in the potential of these media to convey sexual health information [2,3]. Relationships and sex may feature in young people's exchanges on the web, and some see social media platforms as potentially important in supporting positive sexual health outcomes in their communities [4,5]. Nonetheless, in the context of everyday use, young people may not unreservedly accept sexual health content on social media [5-7]. There can also be a disconnect between health promotion conventions and the ways in which young people actually engage on social media [6,8]. To date, some interventions using social media have been shown to improve knowledge of sexually transmitted infection (STI) prevention and to potentially influence sexual behaviors; however, others have resulted in no effect [9-12].

Young people may have concerns about the reputational consequences of associating their social media profile with sexual—or other—health content; they may feel a need to carefully manage how they come across on social media platforms [4,6-8,13,14]. Attention to self-presentation and privacy are required to avoid rumors [7] and negative *drama*, including gossip and bullying [4,6]. Sexual health is potentially stigmatizing, and the avoidance of stigma is a strong factor shaping young people's views [4-6,8]. A recent UK study of 16- to 19-year-olds suggested that the possibility of judgmental reactions from peers toward an overt link to sexual health content does dissuade young people from accessing such content [15].

Young people are aware of not only peers' possible reactions but also those of parents, relatives, and other adults who may be *followers* [4,7,16]. They may consequently want to differentiate what specific social media contacts see and the difficulty of doing so—the concept of *context collapse* [17]—may further explain reluctance toward an association with sexual health content [8,16].

Certain factors can make it more acceptable for young people to access sexual health information or services using social media. These include receiving sexual health messages in an anonymous and unattributable manner and the use of youth-generated and humorous content [6,8,14], although the latter can sometimes miss the mark [14]. Young people also seek credible information from trusted sources [4,5,7,18]. They wish to participate in interventions in a way that allows them to control the degree to which they can be identified and for interventions to be monitored to prevent inappropriate behavior [4,5]. Formats such as question-and-answer forums or private messaging are appealing because of their potential anonymity and privacy [5,7,15,19].

Young people may talk about relationships, sex, and sexual health with each other both on social media and offline, although they may also question the trustworthiness, openness, and confidentiality of peer discussions and favor informal talks with close friends [4-6]. On social media, young people identify a role for friends in endorsing a sexual health intervention or service [5,7]. They may value peer role models in interventions [4,5], and peers' real-life accounts—or fictionalized, plausible scenarios—as potentially effective intervention content [5,6]. Some young people wish to interact with each other as part of web-based, network-based sexual health promotion and knowing their contacts in real-life may be an important design criterion [5].

Although social media have become ubiquitous in health promotion [9-12], few interventions involve messaging instigated by young people themselves, and the knowledge of whether this would work in a school setting is lacking. This gap has held back innovation in school-based sexual health interventions, particularly with respect to informal and social norm–focused approaches that augment classroom learning. This study presents data from a feasibility study of an innovative intervention that employed social media in school-based, peer-led sexual health promotion [20]. We explored students' views to draw out the opportunities and challenges related to the intervention's use of social media and to contribute to wider debate about the use of social media in peer-led, youth-targeted sexual health promotion.

Methods

The Sexually Transmitted Infections and Sexual Health Intervention

The STASH (Sexually Transmitted infections and Sexual Health) intervention (Trial ID: ISRCTN97369178) was adapted from the effective peer-led antismoking ASSIST (A Stop Smoking in Schools Trial) intervention, premised on the diffusion of innovation theory [21,22]. Key adaptions for STASH were the focus on sexual health knowledge, beliefs, and behaviors (rather than on smoking); age group of participants (14-16 years in STASH; 12-13 years in ASSIST); and the use of social media to spread messages alongside offline conversations (ASSIST used conversations only) [20]. The aim was to reduce the risk of STI transmission and promote sexual health. The intervention recruited, trained, and supported students nominated by their friends to serve as peer supporters. During a 10-week intervention period, peer supporters were asked to share sexual health messages from the STASH website among their friends via Facebook (a private group function) and face-to-face conversations. Peer supporters also distributed the URL and password to access the STASH website (printed on cards) to their friends. The STASH website contains curated and bespoke relationships and sex education (RSE) content

(memes, infographics, and links to other sites) and was co-designed with young people and health professionals as part of the intervention development.

Peer supporters invited friends to a private Facebook group in which they posted content from the STASH website (full details of the intervention and study is given in Forsyth et al [20]). The groups were monitored by STASH trainers (youth workers specializing in peer education), who were on hand to encourage the peer supporters, support with tricky questions, and guard against inappropriate posting. Peer supporters were also encouraged to chat to their friends about what they had learned in training and what they liked on the STASH website. The STASH intervention team opted to use Facebook because it was the only platform that offered the option of closed and monitored groups alongside direct sharing of content from the website (via a bespoke application programming interface) and log-in via the platform (without the need for personal information such as a telephone number).

STASH was implemented as a feasibility trial; full evaluation findings are reported elsewhere [23]. The intervention was delivered to 6 schools in central Scotland to assess the feasibility and acceptability of the approach, ahead of a full-scale evaluation. Here, we draw on process evaluation data sources that shed light on the role of social media within the overall intervention approach. The process evaluation sought to assess implementation, mechanisms of change, and context of the intervention via measurement of fidelity, acceptability, exposure, and reach [24]. In this paper, we focus on context, fidelity, and acceptability, specifically in relation to the social media component.

We use *peer supporters* and *friends* to distinguish between those who posted RSE messages (the former) and those who received messages (the latter) and *students* to refer to participating young people collectively.

Setting and Participants

Paired and Group Interviews

We conducted 11 semistructured interviews with 42 students involved in STASH as part of the intervention's process evaluation [20]; 6 paired or group interviews with peer supporters and 5 paired or group interviews with friends. The interviews were conducted with peer supporters (n=20; 9 young women) willing to participate on the day and friends they had invited to contribute (n=22; 10 young women; coordinated with the assistance of the STASH contact teacher). Of the 8 groups and 3 pairs of students, 6 were of single gender (3 all-young women and 3 all-young men) and 5 were of mixed gender. The interviews probed participants' experiences and views of the STASH intervention. For the qualitative work, gender was based on self-identification at school (information on whether students were cis or trans was not collected).

We conducted interviews during the final 2 weeks of the intervention (December 2017). As a result of the constraints of the school setting and working within a single school period, interviews were relatively short (15-30 min). Participant information sheets were circulated to students and their parents or carers (along with opt-out consent forms), and opt-in consent was obtained from all interviewees on the day. All interviews were conducted by female researchers (including KM and CP).

Interviews were audio recorded and transcribed verbatim by a professional service. We carried out a thematic analysis of the qualitative data [25,26], using NVivo 11/12 software (QSR International) to facilitate data management. CP (lead for the process evaluation) coded the entire data set for the process evaluation and MH, using a separate coding frame, coded portions of the data set pertaining to the social media component for this paper. The analytic process for the process evaluation involved data familiarization, that is, summarizing and writing annotated memos; generation of a coding frame that married inductive observation with deductive attention to the study questions of interest; and review and refinement of codes (via discussion between KM, CP, and MH) [25,26]. Key themes were identified via synthesis of coding and discussion between analysts and with reference to the existing literature. The final themes presented here crystallized across the analysis stages and iterations of the write-up.

Quantitative Data

To provide context to, and occasionally quantify, the qualitative themes, we included some descriptive statistics from quantitative data sources across the feasibility study: a web-based questionnaire completed by peer supporters at their final follow-up session (88/104, 84.6% response rate) and a student web-based baseline questionnaire (680/831, 81.8% response rate) and a follow-up questionnaire (603/744, 81.0% response rate, approximately 6 months later), administered to the intervention year group in all 6 schools. Facebook activity (counts of messages, likes, and replies) was extracted via the STASH trainer Facebook account with peer supporters' consent (the trainer was a member of all groups). We also used Facebook Analytics reports (via an application programming interface between the STASH website and Facebook). Consistent with small sample size, exploratory analysis and minor scene-setting role, percentages are rounded to one decimal place, and confidence intervals are not calculated. Methods are not described in detail here, again because of their minor, scene-setting role. They can be found in detail elsewhere [20,23].

Ethical approval for the STASH study was granted by the University of Glasgow MVLS Ethics Committee (project number 20160002).

Results

Talking About the STASH Project and Sexual Health With a Researcher

Field notes taken during interviews record moments of silence and of youth whispering, mumbling, or giggling, and intermittently covering their faces. There was embarrassment when asked to recall message content. This discomfort was evident even among those who said sexual health content was not particularly unusual in social media exchanges; it was evident in those who had noted and read the messages and those who had ignored them.

Variable Engagement With Facebook

Facebook monitoring data showed that across schools, an average of 83.6% (87/104) of peer supporters created a private Facebook group for STASH (8/19, 42% in one school in which Facebook was less commonly used and 79/85, 93% across the other 5 schools). The peer supporters invited an average of 12 friends (including on average 7 non-peer supporters friends) to their group, and they posted an average of 15 messages.

In the follow-up questionnaire, most students said they had a Facebook account (382/452, 84.5%), of whom 58.4% (218/373) said they looked at it regularly. However, 30% (21/70) of peer supporters reported to the peer supporter questionnaire that "the people I wanted to join my group hardly ever use Facebook." This ranged from 0% in one school to 83% in another. It was reported more commonly among young men (27/70, 39%) than young women (17/70, 24%).

In interviews, not all the friends recalled the invitation to join a private Facebook group or the posts that were shared within them. Some friends said that they had not been added to the groups. Others described themselves as nonusers or infrequent users of Facebook or did not have notifications switched on. Taken together, these data suggested that using Facebook may have constrained the spread of messages across the school network.

Friends who did join found the groups acceptable; in the follow-up questionnaire, 60.2% (91/151) of students who were invited to STASH Facebook groups said they were happy to be a member and 35.0% (55/157) said they learned about sexual health by being part of the group.

Private STASH Facebook Groups Were Reassuring but Ran Parallel to Routine Social Media Engagement

In the baseline survey, 68.0% (451/663) of students said that social media were "really important to [their] social life." In interviews, students described using social media (eg, Snapchat and Instagram) routinely to connect with friends and, for some, that was the only purpose. They did not necessarily perceive a distinction between social media and offline interaction, with conversations threading through both modes of communication. Regular social media presence was common, and falling behind mattered, because "[...] then you don't know the gossip the next [...] day" (friend, female), and it could mean that you "[...] miss out on a lot [...]" (friend, male). In the baseline survey, 15.8% (104/659) of students said they "often feel left out by what's happening on social media" (68/364, 18.7% of young women; 35/282, 12.4% of young men).

Among peer supporters, a handful distinguished between their *own*, *actual*, or *real* Facebook use and their STASH group activity. Some said they would have reservations about posting STASH content openly on their personal social media, particularly where family members had been *friended*: "You've got your mum and dad and family on there. That's weird" (peer supporter, female).

For many peer supporters, the fact that STASH involved posting messages to a private group whose membership they controlled, put them at ease with sharing sexual health content on social media:

> *...other people could, like, if you wanted to add them [friends from their year group] to the group or whatever, you could, like, use it, and then it was also, like, private, so, like, if you didn't want anyone to see, they didn't have to see it.* [Peer supporter, female]

> *Yeah. Like if there's certain people you felt uncomfortable sharing that kind of information with, you didn't have to add them.* [Peer supporter, female 2]

This separation of STASH activity from routine social media use was not universal: friends of peer supporters recounted how one peer supporter had promoted STASH on his or her personal profile (outside of the STASH private group), suggesting some variation in the ways that peer supporters combined STASH and routine activity.

Even with the private group option, 2 peer supporters spoke of initial disquiet around inviting friends to join:

> *It's a bit awkward, adding folk, 'cause you don't know what they're going to think of it.* [Peer supporter, female]

However, the apprehension about using Facebook seemed to largely dissipate, with others reflecting that, "it was actually alright" (peer supporter, female). The Facebook group format meant that friends could learn from posts without having to interact with anyone, and this was appreciated by friends: "[the group format meant that you weren't] forced to message them if you know what I mean" (friend, female).

Having a trainer in the group did not appear to worry peer supporters or their friends. Indeed, 59% (42/71) of peer supporters said they were glad to have the trainer in their group (notably more young women than young men: 74% [31/42] vs 36% [10/28]; peer supporter questionnaire). In one group, some peer supporters felt that the trainer presence may have held back willingness to engage with the content and ask questions, whereas others suggested that students were more likely to be hindered by each other.

The Offline STASH Context Legitimated and Augmented Web-Based Posts

Students in the intervention year group learned about STASH either via conversations with peer supporter triggered by the STASH training or via school bulletins and assemblies announcing the project. Some peer supporters recalled that their absence from school to attend training sparked curiosity among their friends and provided opportunities to discuss STASH. Comparison of baseline and follow-up questionnaire data suggested evidence of an increase in some STASH topics from pre to post intervention. For example, student reports of conversations with friends about STIs rose by 15.1% among peer supporters (from baseline of 26/96, 27%); 6.2% among students exposed to at least 1 STASH activity (from baseline of 35/237, 14.8%), and 1.8% among students who reported no exposure (from baseline of 45/214, 21.0%).

Friends of peer supporters who recalled seeing posts often described some initial surprise and subsequent contextualizing of the post as part of the STASH intervention, either by recalling previous conversations or via further discussions with peer supporters: "Yeah, once I saw [the Peer Supporters], they told us what [the invitation] was, obviously, and then that was it" (friend, male). Prior conversations tended to alert friends rather than inform them:

> *Interviewer: Did you – did you know what [the invitation to join the Facebook group] meant at that time?*
>
> *Friend (male): I had an idea, but I didn't know what it would be – [what] it'd be used for.*

The personal connection mattered. The first post sent by peer supporters (which explained the purpose of the STASH groups) caught the attention of a friend in one interview, not because of the content itself but because it was a *serious* post sent by a peer supporter who typically shared humorous content.

The option in STASH to discuss content with message senders appeared to reduce both the unexpectedness and peculiarity of sexual health posts:

> *Interviewer: So then is it a bit weird then to suddenly get a sexual health education thing coming into [your routine social media activity]? Is that quite unusual?*
>
> *Friend (female): Not really. They just put it in the chat and we talk about it.*

Prior offline conversations also gave peer supporters confidence. In one group, a friend recalled conversations before one peer supporter first posted in her STASH Facebook group:

> *she was a bit wary of posting some things but we just kinda said, "Well you might as well, 'cause it's not like [other students are] gonna attack you for it. You were asked to so....* [Friend, female]

From the perspective of these friends, the fact of having been asked legitimated the sending of messages and protected the peer supporter from a negative reaction from peers.

Among peer supporters, the training also helped sensitize them to web-based sexual health content more generally:

> *[...] 'cause I've been told about [sexual health] in an actual training session, I'd probably be more likely to read [a sexual health message] now. Be more interested.* [Peer supporter, male]

Recipients of Messages Are Sometimes Receptive but Do Not Necessarily Engage Visibly on Social Media

Although some friends expressed disinterest in the STASH posts or looked at them out of boredom, others responded with openness and interest. One friend felt that the posts demanded attention because they were different from the usual content:

> *Friend (male): It was weird, 'cause I don't really – you don't really see much o' that (uhuh) posted on Facebook. (Uhuh). It was different to see.*
>
> *Interviewer: So it was a bit different? Was it different good, or different bad, or different, just different?*
>
> *Friend (male): It made you read it, 'cause you don't really see that.*

Even when they were receptive to messages, friends rarely commented substantively on posts. According to the Facebook monitoring data, there were reactions (likes; comments; shares) to just over half of the posts. This was partly about not knowing how to respond or what to say, "[I] don't know what to comment on that" (friend, female).

Similarly, although peer supporters were encouraged to modify the curated content or create their own messages, most peer supporters (50/88, 57%) said they preferred not to do this (peer supporter questionnaire) and only 2 individuals spoke of doing so in interviews. One described the need to alter a message to make it "more grown up" (peer supporter, female) and another to make a message stand out more:

> *...I felt as if someone was scrolling through their Facebook, it wouldn't really stand out, people would just skip through it. So, I took a more creative approach and made it more interesting. Gave it a picture as well....* [Peer supporter, male]

The peer supporters were also proactive in overcoming technical issues, such as taking screenshots and sharing, when the application programming interface's copy/share function failed.

Preferences for Message Design and Content

Friends who were members of multiple groups were able to comment on the differences between them. During one interview, 2 friends noted that in one of their Facebook groups, content was conveyed using a humorous tone, but delivery was nonetheless direct, whereas in another group, messages were more discrete. Both were in mixed-gender Facebook groups, and they speculated that exchanges in single-gender groups may have been more open, though "[...] at the same time you want it to be mixed so you can see both sides" (friend, female). The instant notification feature of Facebook served to remind peer supporters to post regularly. It was also often seen as beneficial by friends; however, those in multiple Facebook groups sometimes viewed multiple notifications as annoying and subsequently muted them.

Participants commonly appreciated humor. They felt it helped draw attention and get messages across. Describing a meme of a nude cat, one participant said:

> *Yes, it was quite funny. It was something to make me read it...* [Friend, male]

Regarding web content in general, another said, "It made it funny but like...you got the information across" (friend, female). Bold colors were also viewed as attention grabbing:

> *...when you're scrolling through and you just see like this big bright thing and you go, 'Oh, what's that'?* [Friend, female]

Brief, clear text and memes or pictures were viewed as "more interesting than reading like a chart or something like that..." (friend, male). Finally, balanced arguments were viewed positively:

> *And it was, like, giving you good and bad points, the ones that makes you, like, read more, that was more informative.* [Friend, male]

Preferences for Social Media Versus Offline Interactions

Preferences regarding social media versus face-to-face communication varied among interviewees. One student explained that web-based interaction was less awkward than face-to-face, unless the chat was with a friend:

> *[...] when it's with [...] one of your best friends it's fine 'cause it's just funny.* [Friend, female]

In another interview, a student said that where others were unknown, web-based interaction was easier and less personal than face-to-face:

> *I don't really mind it too much [being in a group with people you didn't know], it's not like – like actually, talking to them [...].* [Friend, male]

Other friends favored face-to-face conversation, seeing it as less antisocial ("...nobody's talking 'cause they're on their phones'" [friend, female]), whereas others preferred a mix of social media and face-to-face interaction.

A couple of male friends said they would feel uncomfortable talking about sex on either channel but nonetheless recognized that social media posts could trigger discussion, especially when viewed by multiple members of the same friendship network:

> *It's just 'cause social media's a thing that a lot of people use [...]...if it's in a close group of people and if you all see the same post, you might end up like having a conversation, talking about it. If [...] the conversation comes up, and then it can start other questions, and then...So, it can be a good thing.* [Friend, male]

There were mixed views among peer supporters, with some seeing social media as "a good way to get across to other people in our year" (peer supporter, female) and others preferring to focus on offline conversations and awareness raising around the website. These mixed views were reflected in the peer supporter questionnaire, where 42% (36/86) said they preferred to talk, 30% (26/86) preferred to send messages, and 28% (24/86) were indifferent.

Discussion

Principal Findings

The STASH study was innovative in using student-led social media messaging alongside offline conversations in a peer-led, school-based sexual health intervention. We undertook a rigorous and theoretically informed process evaluation. In this study, we explored context, fidelity, and acceptability in relation to social media, to draw out opportunities and challenges of peer-led sharing of sexual health messages. At the time of implementing STASH, Facebook was the only platform to offer private groups, direct sharing of content from a website, and log-in via the platform (without the need for personal information such as a telephone number). However, student engagement with Facebook was variable and hindered message sharing where peer supporters, or their friends, were infrequent or nonusers. Invitation-only, monitored Facebook groups alleviated some peer supporter concerns around sharing sexual health content but led to engagement that ran parallel to their routine social media use. The offline context of STASH (trained peer supporters nominated by their friends; face-to-face conversations initiated by peer supporters) appeared to legitimate and augment web-based posts. Despite this and despite being generally receptive to the posts, few friends engaged visibly on social media. Preferences for content design varied; however, humor, color, and text brevity were important. STASH offered flexibility in ways of sharing messages, and there was no clear preference for either social media or offline message sharing among peer supporters and their friends.

Comparison With Previous Work

Previous research has noted the role of social media in young people's everyday communication with friends [4,6-8]; our data support these findings. Facebook use varied among STASH students, suggesting the need for alternative or multiple platforms. Facebook use among teens in the United States is declining and has recently been overtaken by YouTube, Instagram, and Snapchat [27]. The constantly evolving trends in young people's engagement with social media, including the arrival of new platforms (eg, TikTok), underlines the importance of flexible engagement. Students' design preferences confirmed the importance of offering a diversity of content types and allowing youth to shape messages (even if they rarely took this up in practice) [6,7]. Students actively shaped their engagement with STASH content (eg, opting out of notifications or screenshotting messages). This reinforces earlier conclusions on the importance of nuanced and flexible design, allowing prompt responses to changes in audience preferences and behavior [6,28,29].

The existing literature has discussed young people's preoccupation with their image on social media and caution in associating with specific content in front of certain social media contacts, such as parents [5-8,16]. The notion of *context collapse* (audiences from disparate social contexts collapsed into one) has been invoked to help explain this phenomenon and the potential disruption caused by sexual health content to young people's routine social media use [8,16]. In STASH, the private Facebook groups offered peer supporters a means to circumvent *context collapse* by encouraging them to invite only friends with whom they felt comfortable sharing sexual health content. Thus, as an artifact of this design feature, message sharing in STASH ran parallel to—rather than embedded in—routine use and interactions. The monitoring of Facebook groups by trainers, although acceptable to students, may also have suppressed the *usual* interaction. It likely prevented *drama* and bullying but also likely stifled positive visible engagement.

In line with earlier research [6-8], we detected signs of sexual health stigma, both in terms of sexual health as an interview topic and in the form of hesitation on the part of some peer supporters about distributing STASH messages publicly. Stigma may also have been a factor in the limited visible engagement of friends with STASH content in Facebook groups. Previous

research has also indicated that some young people reject sexual health content on social media [6,7,15]. In this study, some peer supporter friends said they did not spend time viewing STASH messages or turned off notifications of new STASH posts. This may have reflected similar rejection on their part.

Notwithstanding the value placed on privacy and the stigma of sexual health, both our findings and the literature [4-8,14,15] suggest that employing social media in sexual health promotion does hold potential. Our results suggest that the offline context (trained influential peers supported by trainers) helped provide legitimacy to peer-led posting, that widely seen social media posts could trigger offline conversations among friends, and that social media could skirt the potential awkwardness of face-to-face conversations with peers. Yet, there was no universal preference for social media over face-to-face communication. Collectively, these findings lend further support to conclusions indicating that young people may not see social media as a viable sole component of effective sexual health interventions, emphasizing the contribution of offline resources and interaction [5,6,8,18].

Strengths and Limitations

The strength of this study is that it draws from a rigorous and theory-driven process evaluation, which uses mixed methods to interrogate mechanisms of change within the intervention and examine feasibility and acceptability. We sought to achieve rigor via a detailed protocol [20], careful attention to possible reporting bias, and cautious and critical interpretation.

The necessity of conducting interviews within school periods imposed time limits that were shorter than ideal and limited the depth of discussion on social media. Field notes attested to a broader discomfort with discussion of sexual matters. This may have prevented students from admitting to a deeper interest in, or engagement with, the STASH messages. Such reluctance may have been amplified by the school setting and the awkwardness of talking about sexual health with an unknown adult in a position of authority. The paired and group interview format may also have influenced individuals' inclination to discuss their views and experiences, particularly if they differed from others in the group.

Recommendations and Conclusions

The STASH study demonstrates that peer-to-peer sharing of sexual health content via social media in school settings is feasible and acceptable. Our findings attest to the importance of multiple communication channels and opportunities beyond social media alone. They also suggest that to encourage buy-in, interventions must offer young people the flexibility and freedom to choose their preferred way of participation. In offering young people autonomy to engage in ways that are authentic to them, interventions must contend with possible tension brought on by adult oversight of activities for safeguarding purposes. To overcome concerns about association with sexual health content, young people may require a valid justification for message sharing; offline intervention activities can support this. In STASH, friends were often primed, received messages from a known and influential member of their social group, and had the opportunity to discuss content offline. This way, the broader intervention context both legitimated and augmented the social media component. Future studies could further explore which offline intervention activities are most effective at facilitating effective social media approaches.

Social media has strong potential; however, the challenges are significant. The fact that young people engage heavily with social media is, by itself, an insufficient justification to use it in sexual health promotion. It will only be effective if employed in ways that are authentic to young people, mindful of their priorities regarding web-based self-presentation and privacy, and credible as an information source [6,15].

Acknowledgments

The authors are indebted to all the young people and professionals who gave their time and expertise at their stakeholder consultations and events. They thank all the students in the study schools who gave their time to participate in the evaluation and the peer supporters, their contact teachers, and senior management teams for their enthusiasm and commitment throughout STASH. They are indebted to Fast Forward and West Lothian Drug and Alcohol Service who co-designed and delivered the STASH intervention, along with the research team, and to their study collaborators Sally Good (Evidence to Impact Ltd) and Yvonne Kerr and Lesley Walker (Healthy Respect, Edinburgh). In addition, they thank Julie Riddell and Susan Patterson for assistance with fieldwork. The study was funded by the National Institute for Health Research Public Health Research Program (14/182/14). The intervention costs were funded by Edinburgh and Lothian Health Foundation (861) and the Scottish Government (14/182/14). KM, LM, CP, LMD, and SS are supported by the United Kingdom Medical Research Council and the Scottish Government Chief Scientist Office at the MRC/CSO Social and Public Health Sciences Unit, University of Glasgow (MC_UU_12017/11, SPHSU11, MC_UU_12017/14, and SPHSU14). SS was also supported by the MRC Strategic Award (MC-PC-13027). Data are available from the authors on request.

The members of The STASH Study Team are Ross Forsyth, Sarah Barry, Rachael Hunter, Mark McCann, Kirsty Wetherall, and Chiara Broccatelli.

MH is a former MSc student at London School of Hygiene and Tropical Medicine.

Authors' Contributions

KM and LM were the co-principal investigators on the STASH study, CP led the process evaluation, and MH reanalyzed the data on social media (originally for an MSc dissertation). MH, KM, and CP wrote the manuscript with substantive contributions from LE, JB, SS, and LMD (all 4 were co-investigators in the STASH study) and LM. All coauthors approved the final manuscript.

Conflicts of Interest

LM is a scientific adviser to Evidence to Impact Ltd (formerly Decipher Impact Ltd).

References

1. Office for National Statistics. Internet access – households and individuals, Great Britain: 2020. Office for National Statistics.: Office for National Statistics; 2020 Aug 07. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2020 [accessed 2021-01-11]
2. Nadarzynski T, Morrison L, Bayley J, Llewellyn C. The role of digital interventions in sexual health. Sex Transm Infect 2017 Jun 08;93(4):234-235. [doi: 10.1136/sextrans-2016-052926] [Medline: 27932599]
3. Bailey J, Mann S, Wayal S, Abraham C, Murray E. Digital media interventions for sexual health promotion-opportunities and challenges. BMJ 2015 Mar 03;350(mar03 8):h1099-h1099. [doi: 10.1136/bmj.h1099] [Medline: 25736806]
4. Veinot TC, Campbell TR, Kruger D, Grodzinski A, Franzen S. Drama and danger: The opportunities and challenges of promoting youth sexual health through online social networks. : AMIA; 2011 Presented at: AMIA Annual Symposium Proceedings; October 22-26,2011; Washington DC p. 1436-1445 URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3243290/#__ffn_sectitle
5. Veinot TC, Campbell TR, Kruger DJ, Grodzinski A. A question of trust: user-centered design requirements for an informatics intervention to promote the sexual health of African-American youth. J Am Med Inform Assoc 2013 Jul 01;20(4):758-765 [FREE Full text] [doi: 10.1136/amiajnl-2012-001361] [Medline: 23512830]
6. Byron P, Albury K, Evers C. "It would be weird to have that on Facebook": young people's use of social media and the risk of sharing sexual health information. Reproductive Health Matters 2013 May 14;21(41):35-44. [doi: 10.1016/S0968-8080(13)41686-5]
7. Ralph LJ, Berglas NF, Schwartz SL, Brindis CD. Finding Teens in TheirSpace: Using Social Networking Sites to Connect Youth to Sexual Health Services. Sex Res Soc Policy 2011 Feb 22;8(1):38-49. [doi: 10.1007/s13178-011-0043-4]
8. Byron P. Troubling expertise: social media and young people's sexual health. Communication Research and Practice 2015 Nov 16;1(4):322-334. [doi: 10.1080/22041451.2015.1110085]
9. Jones K, Eathington P, Baldwin K, Sipsma H. The Impact of Health Education Transmitted Via Social Media or Text Messaging on Adolescent and Young Adult Risky Sexual Behavior. Sexually Transmitted Diseases 2014;41(7):413-419. [doi: 10.1097/olq.0000000000000146]
10. Gabarron E, Wynn R. Use of social media for sexual health promotion: a scoping review. Glob Health Action 2016 Sep 19;9(1):32193 [FREE Full text] [doi: 10.3402/gha.v9.32193] [Medline: 27649758]
11. Swanton R, Allom V, Mullan B. A meta-analysis of the effect of new-media interventions on sexual-health behaviours. Sex Transm Infect 2015 Feb 28;91(1):14-20. [doi: 10.1136/sextrans-2014-051743] [Medline: 25433051]
12. Wadham E, Green C, Debattista J, Somerset S, Sav A. New digital media interventions for sexual health promotion among young people: a systematic review. Sex. Health 2019:-. [doi: 10.1071/sh18127]
13. Fergie G, Hunt K, Hilton S. Social media as a space for support: Young adults' perspectives on producing and consuming user-generated content about diabetes and mental health. Soc Sci Med 2016 Dec;170:46-54 [FREE Full text] [doi: 10.1016/j.socscimed.2016.10.006] [Medline: 27750067]
14. Nadarzynski T, Burton J, Henderson K, Zimmerman D, Hill O, Graham C. Targeted advertisement of chlamydia screening on social media: A mixed-methods analysis. Digit Health 2019 Feb 01;5:2055207619827193 [FREE Full text] [doi: 10.1177/2055207619827193] [Medline: 30746155]
15. Patterson SP, Hilton S, Flowers P, McDaid LM. What are the barriers and challenges faced by adolescents when searching for sexual health information on the internet? Implications for policy and practice from a qualitative study. Sex Transm Infect 2019 Sep 30;95(6):462-467 [FREE Full text] [doi: 10.1136/sextrans-2018-053710] [Medline: 31040251]
16. Witzel TC, Guise A, Nutland W, Bourne A. It Starts With Me: Privacy concerns and stigma in the evaluation of a Facebook health promotion intervention. Sex. Health 2016;13(3):228. [doi: 10.1071/sh15231]
17. Marwick AE, boyd D. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. New Media & Society 2010 Jul 07;13(1):114-133. [doi: 10.1177/1461444810365313]
18. Evers C, Albury K, Byron P, Crawford K. Young people, social media, social network sites and sexual health communication in Australia: "This is funny, you should watch it". International Journal of Communication 2013;7(1):263-280 [FREE Full text]
19. Fergie G, Hunt K, Hilton S. What young people want from health-related online resources: a focus group study. J Youth Stud 2013 Aug;16(5):579-596 [FREE Full text] [doi: 10.1080/13676261.2012.744811] [Medline: 24748849]

20. Forsyth R, Purcell C, Barry S, Simpson S, Hunter R, McDaid L, et al. Peer-led intervention to prevent and reduce STI transmission and improve sexual health in secondary schools (STASH): protocol for a feasibility study. Pilot Feasibility Stud 2018 Nov 29;4(1):180 [FREE Full text] [doi: 10.1186/s40814-018-0354-9] [Medline: 30519482]
21. Campbell R, Starkey F, Holliday J, Audrey S, Bloor M, Parry-Langdon N, et al. An informal school-based peer-led intervention for smoking prevention in adolescence (ASSIST): a cluster randomised trial. The Lancet 2008 May;371(9624):1595-1602. [doi: 10.1016/s0140-6736(08)60692-3]
22. Rogers EM. Diffusion of preventive innovations. Addictive Behaviors 2002 Nov;27(6):989-993. [doi: 10.1016/s0306-4603(02)00300-3]
23. Mitchell KR, Purcell C, Forsyth R, Barry S, Hunter R, Simpson SA, et al. A peer-led intervention to promote sexual health in secondary schools: the STASH feasibility study. Public Health Research 2020;8(15):1-152 [FREE Full text] [doi: 10.3310/phr08150] [Medline: 33252893]
24. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350(mar19 6):h1258-h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
25. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13(1):117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
26. Green J, Thorogood N. Qualitative Methods for Health Research. London: Sage Publications Ltd; 2009.
27. Pew Research Centre. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/pi_2018-05-31_teenstech_0-01/ [accessed 2020-12-08]
28. Pedrana A, Hellard M, Gold J, Ata N, Chang S, Howard S, et al. Queer as F**k: reaching and engaging gay men in sexual health promotion through social networking sites. J Med Internet Res 2013 Feb 07;15(2):e25 [FREE Full text] [doi: 10.2196/jmir.2334] [Medline: 23391459]
29. Nguyen P, Gold J, Pedrana A, Chang S, Howard S, Ilic O, et al. Sexual health promotion on social networking sites: a process evaluation of The FaceSpace Project. J Adolesc Health 2013 Jul;53(1):98-104. [doi: 10.1016/j.jadohealth.2013.02.007] [Medline: 23583509]

Abbreviations

ASSIST: A Stop Smoking in Schools Trial
RSE: relationships and sex education
STASH: Sexually Transmitted infections And Sexual Health
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 01.06.20; peer-reviewed by P Byron, J Rusow, K Lorimer; comments to author 09.07.20; revised version received 13.08.20; accepted 16.09.20; published 16.02.21.

Please cite as:
Hirvonen M, Purcell C, Elliott L, Bailey JV, Simpson SA, McDaid L, Moore L, Mitchell KR, The STASH Study Team
Peer-to-Peer Sharing of Social Media Messages on Sexual Health in a School-Based Intervention: Opportunities and Challenges Identified in the STASH Feasibility Trial
J Med Internet Res 2021;23(2):e20898
URL: http://www.jmir.org/2021/2/e20898/
doi:10.2196/20898
PMID:33591287

Original Paper

Demystifying the Dark Web Opioid Trade: Content Analysis on Anonymous Market Listings and Forum Posts

Zhengyi Li[1*], MSc; Xiangyu Du[1*], MSc; Xiaojing Liao[1], PhD; Xiaoqian Jiang[2], PhD; Tiffany Champagne-Langabeer[2], PhD

[1]Department of Computer Science, Indiana University Bloomington, Bloomington, IN, United States

[2]The University of Texas Health Science Center at Houston, Houston, TX, United States

[*]these authors contributed equally

Corresponding Author:
Xiaojing Liao, PhD
Department of Computer Science
Indiana University Bloomington
700 N Woodlawn Ave
Bloomington, IN
United States
Phone: 1 8646508137
Email: xliao@indiana.edu

Abstract

Background: Opioid use disorder presents a public health issue afflicting millions across the globe. There is a pressing need to understand the opioid supply chain to gain new insights into the mitigation of opioid use and effectively combat the opioid crisis. The role of anonymous online marketplaces and forums that resemble eBay or Amazon, where anyone can post, browse, and purchase opioid commodities, has become increasingly important in opioid trading. Therefore, a greater understanding of anonymous markets and forums may enable public health officials and other stakeholders to comprehend the scope of the crisis. However, to the best of our knowledge, no large-scale study, which may cross multiple anonymous marketplaces and is cross-sectional, has been conducted to profile the opioid supply chain and unveil characteristics of opioid suppliers, commodities, and transactions.

Objective: We aimed to profile the opioid supply chain in anonymous markets and forums via a large-scale, longitudinal measurement study on anonymous market listings and posts. Toward this, we propose a series of techniques to collect data; identify opioid jargon terms used in the anonymous marketplaces and forums; and profile the opioid commodities, suppliers, and transactions.

Methods: We first conducted a whole-site crawl of anonymous online marketplaces and forums to solicit data. We then developed a suite of opioid domain–specific text mining techniques (eg, opioid jargon detection and opioid trading information retrieval) to recognize information relevant to opioid trading activities (eg, commodities, price, shipping information, and suppliers). Subsequently, we conducted a comprehensive, large-scale, longitudinal study to demystify opioid trading activities in anonymous markets and forums.

Results: A total of 248,359 listings from 10 anonymous online marketplaces and 1,138,961 traces (ie, threads of posts) from 6 underground forums were collected. Among them, we identified 28,106 opioid product listings and 13,508 opioid-related promotional and review forum traces from 5147 unique opioid suppliers' IDs and 2778 unique opioid buyers' IDs. Our study characterized opioid suppliers (eg, activeness and cross-market activities), commodities (eg, popular items and their evolution), and transactions (eg, origins and shipping destination) in anonymous marketplaces and forums, which enabled a greater understanding of the underground trading activities involved in international opioid supply and demand.

Conclusions: The results provide insight into opioid trading in the anonymous markets and forums and may prove an effective mitigation data point for illuminating the opioid supply chain.

(J Med Internet Res 2021;23(2):e24486) doi:10.2196/24486

KEYWORDS

opioids; black market; anonymous markets and forums; opioid supply chain; text mining; machine learning; opioid crisis; opioid epidemic; drug abuse

Introduction

Background

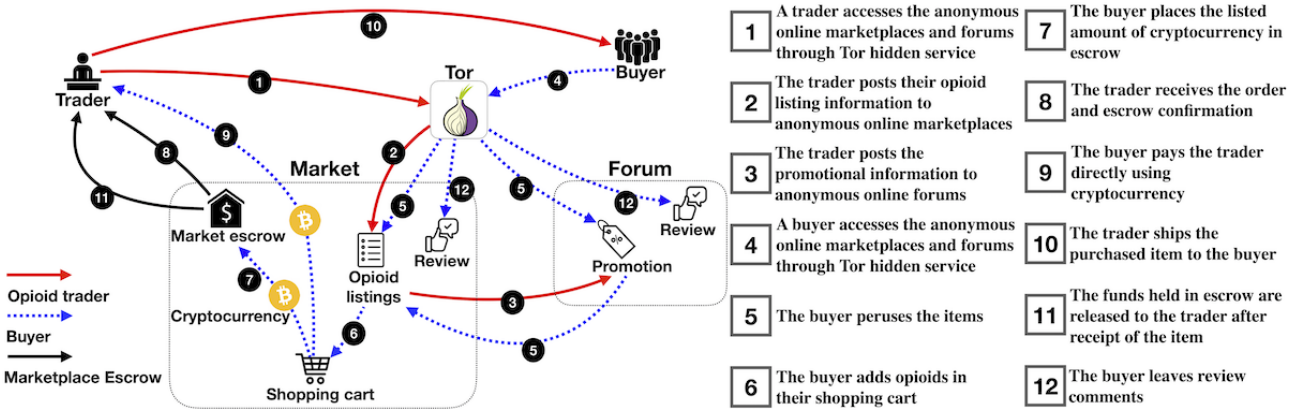
Overdoses from opioids, a class of drugs that includes both prescription pain relievers and illegal narcotics, account for more deaths in the United States than traffic deaths or suicides. Overdose deaths involving heroin began increasing in 2000 with a dramatic change in pace, and as of 2014, 61% of drug overdoses involved some type of opioid, inclusive of heroin [1]. Deaths involving fentanyl nearly doubled from the previous year's rate in 2014, 2015, and 2016 [2]. To reduce opioid-related mortality, there is a pressing need to understand the supply and demand for the product; however, no prior research that provides a greater understanding of the international opioid supply chain has been conducted.

The past 10 years have witnessed a spree of anonymous online marketplaces and forums, mostly catering to drugs in anonymous ways and resembling eBay or Amazon. For instance, SilkRoad, the first modern darknet market and best known as a platform for selling illegal drugs, was launched in February 2011 and subsequently shut down in October 2013 [3]. However, its closure catalyzed the development of multiple other anonymous marketplaces. Compared with traditional opioid supply methods [4], the role of anonymous online marketplaces and forums has become more important because of its stealthiness and anonymity: using this type of virtual exchange, anyone can post and browse the opioid product listings, regardless of their technical background. It raises new challenges for new law enforcement agencies to identify opioid suppliers, buyers, or even takedown the marketplace. Further compounding the issue from a law enforcement perspective, it is nontrivial to obtain complete opioid listings from the darknet markets, interpret the jargon used in the darknet forum, and holistically profile opioid trading and supplying activities.

Underground Opioid Trading

Anonymous online marketplaces are usually platforms for sellers and buyers to conduct transactions in a virtual environment. They usually come with anonymous forums for sellers and buyers to share information, promote their products, leave feedback, and share experiences about purchases. To understand how it works, we describe an opioid transaction's operational steps on the anonymous online marketplaces and forums. We present a view about how such services operate and how different entities interact with each other (Figure 1).

Figure 1. Overview of the opioid trading in the anonymous marketplaces and forums.



First, an opioid trader, who intends to list the selling information and find potential customers, will first access the anonymous online marketplaces and forums, using an anonymous browsing tool such as a Tor client or a web-to-Tor proxy (step 1 in Figure 1) [5,6]. Anonymous online marketplaces and forums usually operate as hidden Tor services, which can only be resolved through Tor (an anonymity network). Once connected to the anonymous online marketplaces (eg, The Empire Market and Darkbay), the opioid trader will create an account as a seller and post their opioid listing information (including product, price, origin country, an acceptable shipping destination, payment method, quantities left, shipping options—shipping days or shipping companies, and refund policy; step 2). Figures 2 and 3 illustrate examples of opioid listings in The Versus Project and Alphabay. The opioid trader will also use an anonymous online forum (eg, The Hub Forum) to post promotional information to attract potential customers (step 3).

Suppose that an opioid buyer wants to purchase opioids. The opioid buyer (client) will also access the anonymous online market and create an account in each anonymous marketplace before they can find the listings of opioids (step 4). After perusing the items available on the anonymous online market (step 5), the buyer will add opioids to their shopping cart (step 6). When the buyer wants to check out and make a purchase using cryptocurrency (eg, Bitcoin), if the trader accepts payment through an anonymous online marketplace as an escrow, the buyer will place the listed amount of cryptocurrency in escrow (step 7). Then, the trader receives the order and escrow confirmation (step 8). Otherwise, the buyer will pay the trader directly using cryptocurrency or any other payment method accepted by the trader (step 9) [7]. Note that the escrow mechanism is widely deployed in the anonymous online market because it helps to build trust and resolve disputes between sellers and buyers. When the purchase is made, the opioid trader

ships the purchased item to the buyer (step 10). Once the item is received, the buyer finalizes the purchase by notifying the anonymous online marketplace to release the funds held in escrow (step 11) [8,9]. After that, an opioid buyer often leaves review comments under the product listing or discusses the purchasing experience in the forum (step 12).

Figure 2. Example of opioid listings in The Versus Project.

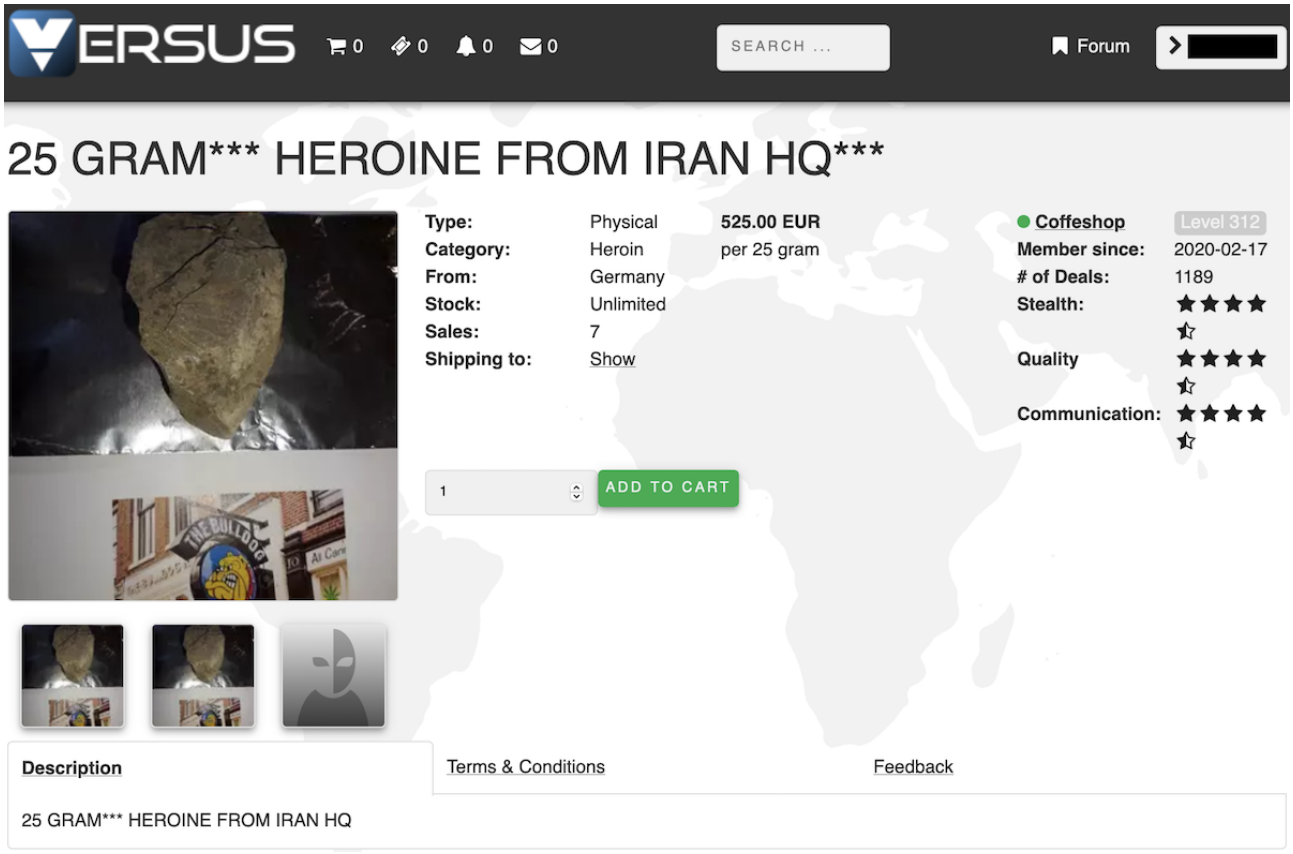
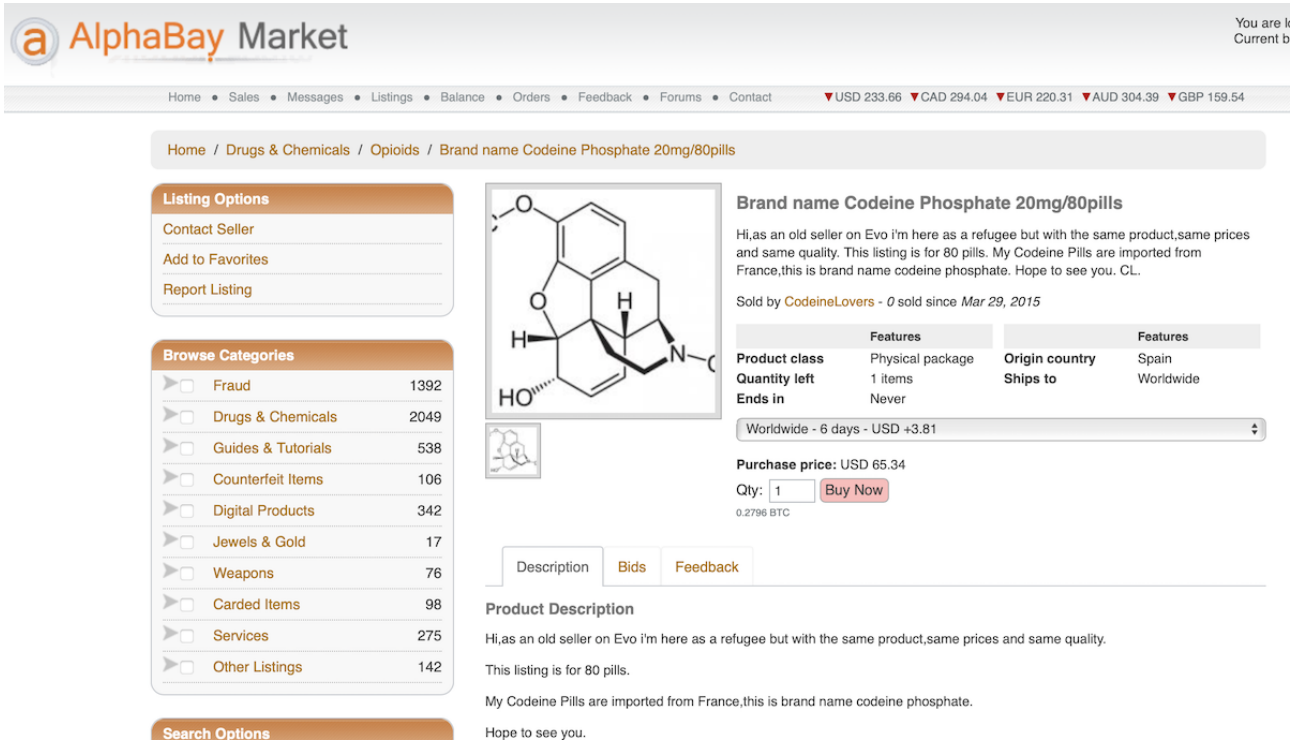


Figure 3. Example of opioid listings in Alphabay.



Prior Work

Recent years have witnessed the trend of studying opioid use disorders using anonymous marketplaces and forums data [5,8,10] and public social media data (eg, Twitter and Instagram) [11-14]. Gilbert et al [15] described changes in the conceptualizations, techniques, and structures of opioid supply chains and illustrated the diversity of transactions beyond the traditionally linear conceptualizations of cartel-based distribution models. Quintana et al [16] and Fernando et al [17] presented the results of the international drug testing service for opioid commodities from the anonymous marketplaces and showed that most opioid substances contained the advertised ingredient and most samples were of high purity. Dasgupta et al [18] collected opioid listings on Silk Road to analyze the prices of diverted prescription opioids. Duxbury et al [6] evaluated the role of trust in online drug markets by applying exponential random graph modelling to underground marketplace transactions. The results show that vendors' trustworthiness is a better predictor of vendor selection than product diversity or affordability. Considering social media data (eg, Twitter and Instagram), Nasralah et al [14] proposed a text mining framework to collect opioid data from social media and analyzed the most discussed topics to profile the opioid epidemic and crisis. Mackey et al [13] collected tweets related to the opioid topic to identify illicit online pharmacies and study the illegal sale of opioids in online marketing. Cherian et al [12] gathered codeine misuse data from Instagram posts to understand how misuse is happening and its misused form. Recently, Balsamo et al [11] used a language model to expand vocabularies for opioid substances, routes of administration, and drug tampering on Reddit data from 2014 to 2018 and investigated some important consumption-related aspects of the nonmedical abuse of opioid substances. However, to the best of our knowledge, no large-scale study, which may cross multiple anonymous marketplaces and is cross-sectional, has been conducted to profile the opioid supply chain and unveil characteristics of opioid suppliers, commodities, and transactions.

Goals

This paper seeks to complement current studies widening the understanding of opioid supply chains in underground marketplaces using comprehensive, large-scale, longitudinal anonymous marketplace and forum data. To this end, we propose a series of techniques to collect data; identify opioid jargon terms used in the anonymous marketplaces and forums; and profile the opioid commodities, suppliers, and transactions. Specifically, we first conducted a whole-site crawl of anonymous online marketplaces and forums to solicit data. We then developed a suite of opioid domain–specific text mining techniques (eg, opioid jargon detection and opioid trading information retrieval) to recognize information relevant to opioid trading activities (eg, commodities, price, shipping information, and suppliers). Subsequently, we conducted a comprehensive, large-scale, longitudinal study to demystify opioid trading activities in anonymous markets and forums.

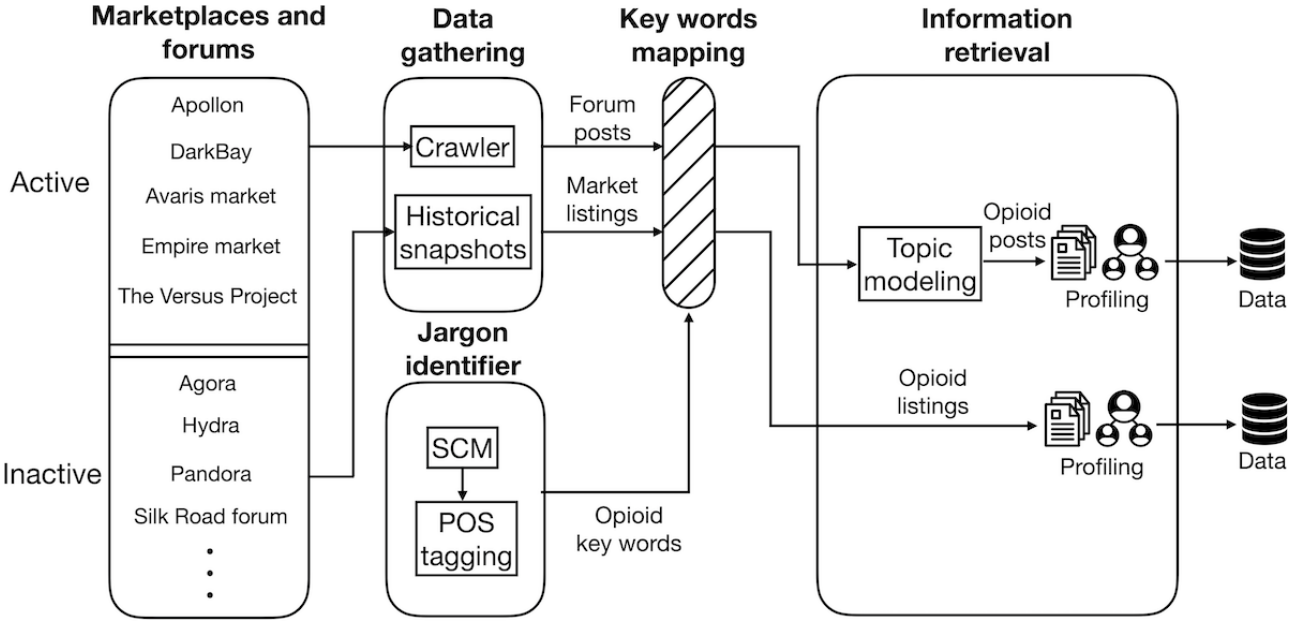
The contributions of this study are elaborated below. First, we designed and implemented an anonymous marketplace data collection and analysis pipeline to gather and identify opioids data in 16 anonymous marketplaces and forums over a period of almost 9 years between 2011 and 2020. Second, we fine-tuned the semantic comparison model proposed by Yuan et al [19] for opioid jargon detection, which can recognize the opioid jargon as innocent-looking terms and the dedicated terms only used in the anonymous marketplaces and forums. In this way, we generated a rich underground marketplace opioid vocabulary of 311 opioid keywords with 13 categories. Third, we conducted a comprehensive, large-scale, longitudinal study to measure and characterize opioid trading in anonymous online marketplaces and forums. Specifically, using a large-scale and cross-sectional data set, we characterized the activeness and cross-market activities of opioid suppliers, investigated popular opioid commodities as well as their evolution and price trends, and outlined a picture of origins and shipping destinations appearing in opioid transactions in anonymous marketplaces and forums. We believe our findings will provide insight into opioid trading in the anonymous markets and forums for law enforcement, policy makers, and invested health care stakeholders to understand the scope of opioid trading activities and may prove an effective mitigation data point for illuminating the opioid supply chain.

Methods

Overview

This section elaborates on the methodology used to identify opioid trading information in the anonymous market and forums. We illustrate the methodology pipeline (Figure 4). Specifically, we collected approximately 248,359 unique listings and 1,138,961 unique forum traces (ie, threads of posts) from 10 anonymous online marketplaces and 6 forums. We then identified 311 opioid keywords and jargons to recognize 28,106 listings and 13,508 forum traces related to underground opioid trading activities. Finally, we used natural language processing techniques to extract opioid trading information to characterize underground opioid commodities, suppliers, and transactions.

Figure 4. Overview of the methodology workflow. SCM: semantic comparison model, POS: part-of-speech.



Data Collection

Our research collected product listings and forum posts from 10 anonymous online market places and 6 forums. Our study determined the underground marketplace and forum list based on darknet site search engines and previous research works [20]. More specifically, we used darknet site search engines (such as Recon, Darknet live, Dark Eye, dark.fail, and DNStats [21,22]) to search underground marketplaces and forums and then manually validated their activeness. In our study, we only selected marketplaces with more than 30 opioid listings. In this way, we gathered 5 active underground marketplaces with opioid listings. Note that some high-profile underground marketplaces and forums are frequently deactivated or have been shut down by law enforcement authorities [23]. Hence, we also gathered snapshots of 5 underground marketplaces and 6 forums collected by the anonymous marketplace archives programs and previous research projects [20].

To collect the listing information of 5 anonymous online marketplaces (ie, Apollon, Avaris, Darkbay, Empire, and The Versus Project), we conducted a whole-site crawl. The crawler was implemented in Python and used the Selenium module to launch browsers and to send crawling requests [24]. To avoid blocking from the marketplace, we provided as an input to the scraper a session cookie that we obtained by manually logging into the marketplace and solving a Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA). However, some sites, namely, the Empire Market, forced users to log out when the life span of the session cookie was expired. In this case, we had to manually repeat the previous process. In addition, we set parameters such as sleeping time to limit the speed of crawling.

In total, we collected 248,359 listings of 10 anonymous online marketplaces between December 2013 and March 2020. For forum corpora, we gathered 1,138,961 traces (spanning from June 2011 to July 2015) from the underground forums The Hub, Silk Road, Black Market, Evolution, Hydra, and Pandora. Table 1 summarizes the data sets used in this study. Note that some forums, such as Pandora and Evolution, were associated with the corresponding marketplaces and mainly served as discussion platforms for marketplace buyers and vendors. In addition, the measurement dates vary across different marketplaces and forums, as they have different life spans.

Table 1. Data set summary of marketplaces and forums that were collected for this study.

Name	Type	Lifetime	Measurement dates	Number of traces/listings	Number of opioid traces/listings
Agora	Marketplace	December 2013 to August 2015	January 2014 to July 2015	140,266	12,051
Alphabay	Marketplace	December 2014 to July 2017	December 2014 to July 2015	21,679	1344
Hydra	Marketplace	March 2014 to November 2014	August 2014 to October 2014	3048	218
Pandora	Marketplace	October 2013 to November 2014	December 2013 to November 2014	20,013	1749
Evolution	Marketplace	January 2014 to March 2015	April 2014 to March 2015	54,196	4954
Apollon	Marketplace	May 2018 to March 2020	September 2018 to February 2020	2921	2552
Empire	Marketplace	February 2018 to August 2020	April 2018 to March 2020	2995	2548
The Versus Project	Marketplace	November 2019 to now	November 2019 to March 2020	233	202
Avaris	Marketplace	October 2019 to August 2020	October 2019 to February 2020	291	286
Darkbay	Marketplace	July 2019 to September 2020	July 2019 to February 2020	2717	2112
Black Market	Forum	December 2013 to February 2014	December 2013 to February 2014	52,127	669
Pandora	Forum	October 2013 to September 2014	January 2014 to September 2014	18,640	798
Hydra	Forum	March 2014 to November 2014	April 2014 to September 2014	887	41
The Hub	Forum	January 2014 to now	January 2014 to July 2015	53,973	1082
Evolution	Forum	January 2014 to March 2015	January 2014 to November 2014	166,641	2682
Silk Road	Forum	January 2011 to November 2014	June 2011 to November 2013	846,693	34,519

Opioid Jargon Identification

Our study used opioid keywords and jargons to recognize listings and forum traces related to underground opioid trading activities. Our opioid jargon identification procedure implemented a modified semantic comparison model [19]. This model employed a neural network–based embedding technique to analyze the semantics of words in different corpora. In particular, in the semantic comparison model, the size of the input layer was doubled while not expanding either the hidden or the output layer. In this way, the same word from 2 different corpora will build separate relations, in terms of weights, from the input to the hidden layer during the training, based on their respective datasets, while ensuring that the contexts of the word in both corpora are combined and jointly contribute to the output of the neural network through the hidden layer. Hence, every word has 2 vectors, each describing the word's relations with other words in one corpus. In the meantime, these 2 vectors are still comparable because they are used together in the neural network to train a single skip-gram model for predicting the surrounding windows of context words.

Our modification of the semantic comparison model will generate comparable word embeddings for opioid jargon words in legitimate documents (ie, benign corpora embedding) and in underground corpora (ie, underground corpora embedding). Specifically, our modification used a series of opioid keywords collected to generate their benign corpora embeddings and then searched for words whose underground corpora embeddings were close to the opioid keywords' benign corpora embeddings. We output the top 100 proper nouns in the underground corpora in our implementation, whose embeddings showed the closest cosine distance to the known opioid keywords.

We trained the semantic comparison model using the traces of Reddit as the benign corpora and the traces of the anonymous marketplaces/forums as the underground corpora. The parameters of the model were set as default [19]. Thus, we identified 58 opioid jargon used in the anonymous marketplaces and forums (Table 2). Combining opioid jargon with known opioid product names [25-27], we generated an opioid keyword data set consisting of 311 opioid keywords with 13 categories. We manually validated all keywords and the corresponding

categories to guarantee their correctness. Our keywords included almost all the vocabularies of opioid substances mentioned in the study by Balsamo et al [11] and further expanded their results with specific medicine codes, product names, and special colloquialisms (eg, M523, Ultram hydrochloride tramadol 200, and H3 brown sugar).

Table 2. Opioid jargons used in the anonymous online marketplaces and forums.

Category	Jargons
Heroin	gunpowder, pearl tar (black pearl tar), speedball, heroin #4, diacetylmorphin, and h3 brown sugar
Fentanyl	chyna (china white), acetylfentanyl (acetyl fentanyl), phenaridine, and duragesic
Buprenorphine	subutex and suboxone
Oxycodone	roxy, roxi, roxies, roxys, oxynorm, A215, K8/K9, M15/30, blueberries, A15, OC30/80, OP80, oxyneo, M523/IP204/C230, bananas, V4812, and CDN 80
Dihydrocodeine	DHC[a]
Oxymorphone	panda and o bomb
Morphine	zomorph, mscontin (ms contin), skenan, oramorph, and kadian
Methadone	amidone, methadose, and chocolate chip cookies
Hydromorphone	hydromorph
Hydrocodone	lortab, norcos, zohydro, IP109/110, and M367
Tramadol	UDT[b] 200
Codeine	thiocodin and lean
Others	tapentadol, tapalee, and nucynta

[a]Dihydrocodeine bitartrate.

[b]Ultram hydrochloride tramadol.

Topic Modeling of Forum Posts

Our goal here was to identify anonymous forum posts with the topics of opioid commodity promotion (eg, listing promotion) and review (eg, report fake opioid vendors). We then analyzed these forum posts to profile underground opioid trading behaviors.

To identify forum posts related to opioid commodity promotion and review, our methodology was designed to filter forum posts with opioid keywords and then use a classifier to the posts with the topics of interest. The classifier was built upon transfer learning and a crafted objective function that heavily weighs the penalty of misclassifying a positive instance.

The model training process for opioid promotion and review posts' detection consists of 3 stages: model initialization, transfer learning, and model refining. First, 2 neural network models with 3 hidden layers are trained on the data sets (Table 3) for model initialization. Then, in the transfer learning stage, the aforementioned models are transferred using manually labeled 800 positive samples (D_p) and 800 negative samples (D_c; Table 3), with the purpose of adjusting the model to fulfill the promotion posts and opioid review detection. Due to the difficulty in collecting the opioid promotion and review posts, the number of positive samples D_p is relatively small compared with negative samples D_c. Note that a sample is only annotated when both of the 2 graduate student annotators agreed with each other. For the data annotation, intercoder reliability measured with Cohen kappa coefficients was 0.74 for promotion post labeling result and was 0.68 for the opioid review labeling result. Considering the imbalance of D_p and D_c, we modified the loss function of the model to make it weigh the penalty of misclassifying a positive instance. The objective function is as follows:

where *LL (z)* is the log loss, that is, 3 log (1 + exp (−z)). C_+ and C_- denote the penalty factors for misclassifying the positive and negative instances, respectively, whereas λ is the regularization coefficient and ||ω|| is the regularization term, which is the L1-norm. In the above objective function, C_+ is always larger than C_-, which means that the penalty of misclassifying a positive instance is larger than that of a negative instance. In general, the correlation between the penalty factors and the number of samples is set as  , where P and C are the sizes of D_p and D_c, respectively.

Table 3. Data sets used in the forum post modeling.

Topic	Positive samples (n)		Negative samples (n)	
	Model initialization data set	Annotated anonymous market/forum data set	Model initialization data set	Annotated anonymous market/forum data set
Promotion	Listing descriptions in the marketplace Agora and Alphabay (60,000)	Listing descriptions and product promotions in the anonymous marketplace (1000)	Amazon review data set [28] (30,000) and Yahoo! Answers data sets [29] (30,000)	Nonpromotion (ie, review and question answering) posts in anonymous markets and forums (1000)
Review	Amazon review data set [28] (100,000)	Review posts in an anonymous marketplace (1000)	Yahoo! Answers data sets [29] (100,000)	Nonreview posts in anonymous markets and forums (1000)

Finally, in the model refining stage, the model is trained for other 2 iterations using the same objective function. We manually investigated the results by randomly sampling 10% of data records during each iteration and adding false positive samples into the unlabeled set. Our model was evaluated via 10-fold cross-validation. The review detection model yielded a mean precision of 81.5% and an average recall of 80.1%, whereas for the promotion detection model, it yielded a mean precision of 88.1% and an average recall of 85.1% (Table 4).

Table 4. The results and 95% CIs of forum posts' topic modeling.

Topic modeling methods	Promotion topic		Review topic	
	Precision	Recall	Precision	Recall
MALLET[a] document classification, mean (SD)				
NaiveBayes	81 (2)	80 (3)	64 (2)	87 (3)
C45	66 (7)	68 (11)	56 (6)	75 (9)
Decision tree	83 (3)	51 (3)	71 (2)	61 (4)
MALLET topic modeling, n (%)				
Unsupervised topic modeling	814 (48)	814 (81.40)	939 (53.69)	939 (93.90)
Our model, mean (SD)	88 (1)	85 (2)	82 (1)	80 (1)
Baseline, mean (SD)	84 (1)	84 (3)	76 (3)	74 (2)

[a]MALLET: Machine Learning for Language Toolkit.

We compared our method with the state-of-the-art topic modeling method Machine Learning for Language Toolkit (MALLET) [30] and our model without transfer learning stage (baseline). Our experiment evaluated MALLET on our annotated anonymous marketplace and forum data set (Table 3) using 3 classification algorithms in the document classification tool (package cc.mallet.classify class in MALLET's JavaDoc API [Application Programming Interface]). In particular, MALLET is retrained and evaluated via 10-fold cross-validation. We also applied the MALLET topic modeling toolkit (package cc.mallet.topics MALLET's class in JavaDoc API) on the same data set to predict the type of topic. The baseline model was applied directly to the labeled data (Table 3) and evaluated using 10-fold cross-validation. We used the metrics of precision and recall to compare the performance of different topic modeling methods. As shown in Table 4, our results indicate that our approach significantly outperforms MALLET and the baseline model in terms of both precision and average recall.

In this way, we collected 7100 promotion posts and 6408 review posts from forum posts in total.

Opioid Trading Information Retrieval

For each marketplace listing and forum posts related to opioid promotion, we extracted 8 properties: vendor name, product, price, number of products sold, advertised origins, acceptable shipping destinations, and whether escrow or not. For the forum posts on the topic of the opioid commodity review, we recognized the sentiment of the review. Below, we elaborate on the methodology used to identify each of the properties:

- Vendor name: To identify the vendor name, we designed a parser to identify the authors of the listings and promotional posts by applying platform-specific heuristics, which we manually derived from each marketplace and forum's HTML templates.
- Product: We recognized the type of opioid in each listing's description content using the opioid keyword data set generated in the previous step.
- Price: We used a price extraction model [31], which was trained on the underground forum corpora, to extract listing price information (Figures 2 and 3). Our study further determined the per-gram price of opioid products by dividing the listing price by the amount of products. More specifically, we designed a set of regular expressions to extract the amount of opioids sold per listing. For instance, in Figure 3, 1.6 g (20 mg × 80 pills) codeine is sold for US $65.34. Note that following previous works [8], we also dismissed the abnormal price that was greater than 5 times

the median of the remaining samples as well as less than 25% of the value of the median.
- Number of products sold: Listings of 5 marketplaces (Alphabay, The Versus Project, Apollon, Empire, and Darkbay) consist of the number of items that have been sold (as shown in Figure 2). Hence, we applied the parser's feature of identifying the number of sold, which we manually derived from each marketplace's HTML templates, if such information can be found in the marketplace.
- Advertised origins and acceptable shipping destinations: We parsed the advertised origins and acceptable shipping destinations from the HTML template of marketplace listings and used the country name dictionary to find the country names from a forum post. We considered the contextual information based on the keywords *ship*, *origin*, and *destination*.
- Whether escrow: In the marketplaces Alphabay, Apollon, and Empire, the product listing usually has a field to indicate whether the escrow is supported. Hence, we designed a parser to obtain this information. In forum posts, we used the keyword *escrow* to match each forum post with the topic of promotion to find out whether the trader accepts *escrow*.
- Review sentiment: To investigate the sentiment of the opioid product reviews, we applied the chi-square score–based sentiment analysis model to classify the product review into positive and negative [32].

To evaluate the aforementioned methods for extracting properties, we randomly chose 1000 listings for each property and manually annotated the properties as ground truth. We evaluated our method on our annotated data set, which yields an accuracy of over 90% for each property extraction, as shown in Table 5.

Table 5. The results of calculating accuracy of opioid information retrieval.

Property	Number of ground truth, n	Accuracy, n (%)
Vendor name	1000	1000 (100)
Product	1000	954 (95.40)
Price	1000	1000 (100)
Number of products sold	1000	1000 (100)
Advertised origins	1000	1000 (100)
Acceptable shipping destinations	1000	1000 (100)
Whether escrow	1000	1000 (100)
Review sentiment	1000	926 (92.60)

Results

Landscape

In total, we collected 248,359 listings from 10 anonymous online marketplaces and 1,138,961 traces (ie, threads of posts) from 6 underground forums. Among them, we identified 28,106 opioid product listings and 13,508 opioid-related promotional and review forum traces from 5147 unique opioid suppliers' IDs and 2778 unique opioid buyers' IDs. As observed in our data set, the top 3 marketplaces with the most opioid listings are Agora, Evolution, and Apollon.

In our study, we found that 23.78% (9896/41,614) listings and traces were identified with the help of 58 opioid jargons (Table 2). Among them, *suboxone* and *subutex* medicines are most frequently mentioned by 2917 times in listings and traces in 10 platforms, followed by roxy series (ie, roxy, roxi, roxies, and roxys) with 2022 times and *Lean* with 1256 times. Both *K9* and *M30* were mostly found in Darkbay, within 384 listings in the year 2020, whereas *Lean* appeared 141 times in Empire listings.

Note that we should not overestimate the number of suppliers and buyers given the number of IDs found in this research, but we regarded it as the upper-bounded number of the opioid suppliers and buyers. This is because the same user could have different IDs, and the same ID in different marketplaces can point to different users. Owing to the anonymity of the underground marketplaces and forums, there exists no ground truth to link users with their IDs.

Characteristics of Commodities

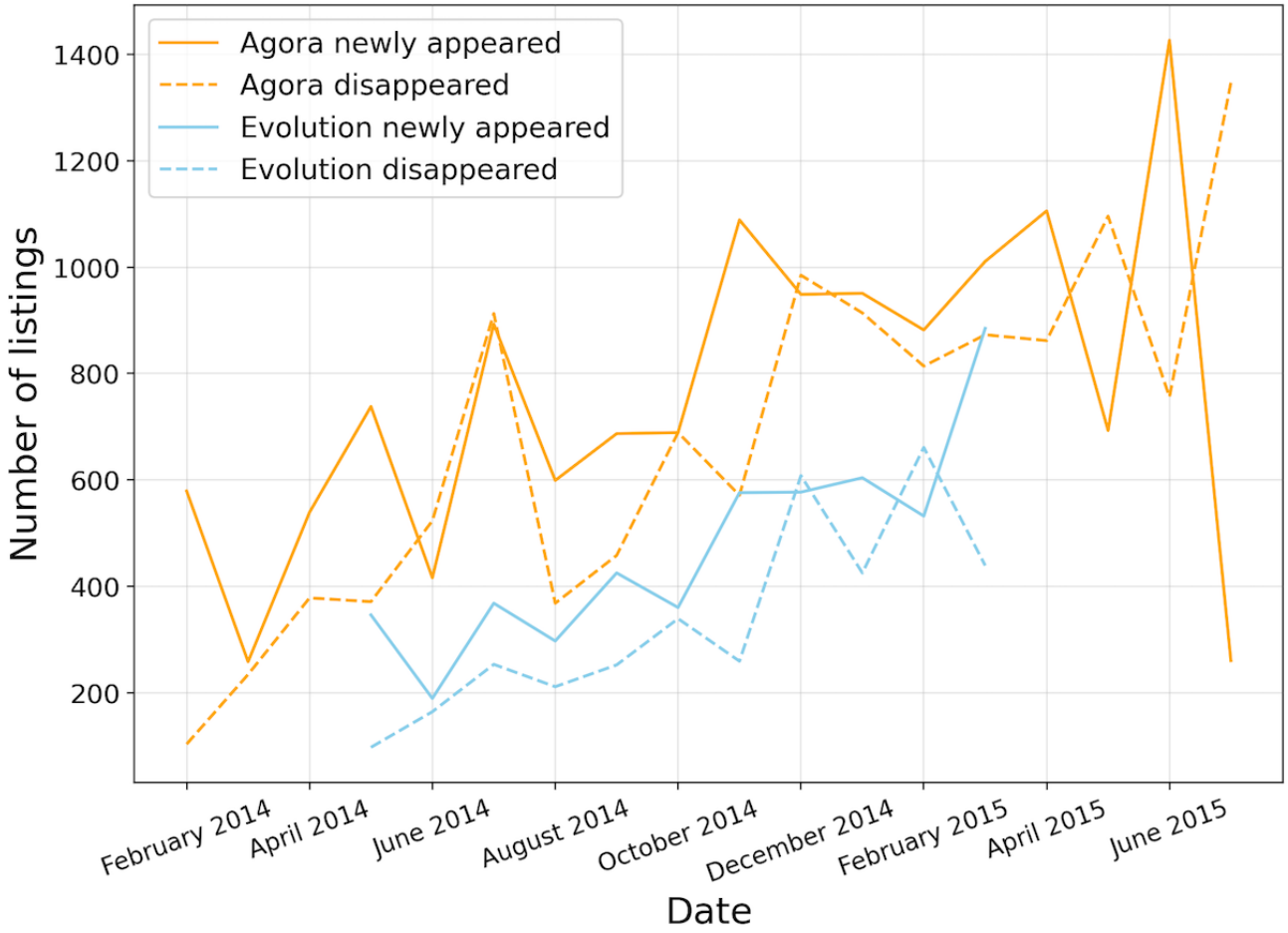
We list the top 5 opioids with most listings and their average prices in 2014, 2015, 2019, and 2020 (Table 6). In general, heroin was found to be the most popular item on the anonymous online market, followed by oxycodone. We also noticed that heroin dropped by 60%, whereas codeine increased by 32% from 2014 to 2019, which is roughly in line with the temporal trend of popularity of opioid substances on Reddit from 2014 to 2018, as mentioned in the study by Balsamo et al [11]. More remarkably, we observed 2011 listings of *China white* (or the slang term *chyna*), a *designer opioid* with significant medical concerns due to its deadly clinical manifestations, in 6 marketplaces and 5 forums. The earliest listing was observed on the SilkRoad in June 2011. Moreover, we notice that most of the top opioids have a lower mean price than their retail prices. For instance, the United Nations Office on Drugs and Crime [33] reported that the average retail price of heroin was US $267 per gram in the United States in 2014 and 2015, which is almost twice as much as the price in anonymous marketplaces (US $130-190 per gram). In addition, we see a trend of drop in price from 2014 to 2019 of some opioids such as heroin, oxycodone, and fentanyl, which may lead to severe overdoses.

Table 6. Popular opioids according to different years. Note that data for 2020 only included data from January to March; price per gram is in US $.

2014			2015			2019			2020		
Name	Number of listings	Price (per gram), mean (SD)	Name	Number of listings	Price (per gram), mean (SD)	Name	Number of listings	Price (per gram), mean (SD)	Name	Number of listings	Price (per gram), mean (SD)
Heroin	4251	129.5 (99.9)	Heroin	2408	185.6 (138.4)	Heroin	1697	73.0 (49.8)	Heroin	611	67.0 (37.8)
Oxycodone	3086	660.3 (445.3)	Oxycodone	2079	1239.1 (843.0)	Oxycodone	1078	520.8 (444.9)	Oxycodone	356	590.6 (450.4)
Fentanyl	1397	1116.4 (647.5)	Fentanyl	1450	1546 (909.4)	Codeine	418	80.3 (61.1)	Fentanyl	149	154.2 (123.2)
Buprenorphine	934	2764.9 (2007.8)	Buprenorphine	571	4243.7 (3006.1)	Tramadol	331	16.1 (11.7)	Buprenorphine	90	2083.7 (1471.7)
Tramadol	839	21.5 (14.4)	Tramadol	555	29.8 (29.1)	Fentanyl	282	247.8 (172.1)	Hydrocodone	89	1183.1 (374.0)

When evaluating the activeness of the underground opioid listing, we measured the monthly *newly appeared* and *disappeared* listings in 2 anonymous online marketplaces: Agora and Evolution. Figure 5 shows the results. We observed that large amounts of opioid listings were newly posted every month in the Agora marketplace, which had a relatively higher rate of increase than the disappeared rate in terms of listings. A similar scenario was observed in the marketplace Evolution. We also observed an increase in newly appeared listings in the Agora marketplace in April 2015. This may be because of the shutdown of the Evolution marketplace in March 2015.

Figure 5. Number of monthly newly-appeared and disappeared listings in the marketplaces Agora and Evolution.



We observed the same listings posted in different marketplaces and illustrate the dependency of the same opioid listings among different marketplaces. Note that we determined if 2 listings are identical by matching the same elements (ie, listing's title

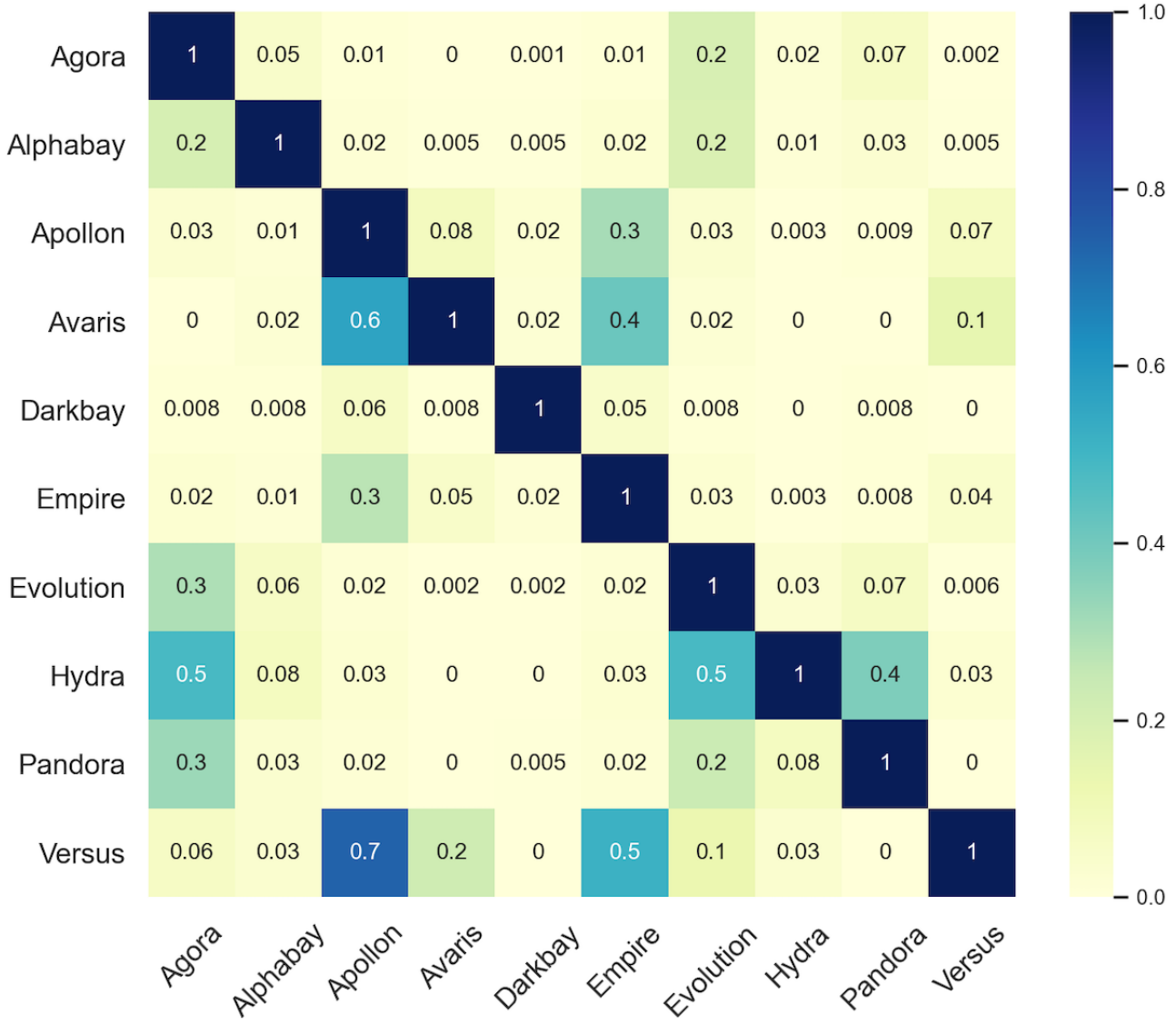
and description information and the vendor's name) in 2 listings. We observed that the marketplaces of Agora and Evolution shared 530 opioid listings from 290 unique supplier IDs. The opioid commodity with most listings across different marketplaces was *#4 White Vietnamese Heroin*, which can be found in the marketplaces Agora, Evolution, Hydra, and Pandora.

Characteristics of Suppliers

To understand the scale of opioid suppliers on the anonymous online market, we scanned the listings of 10 marketplaces and the promotional posts of 6 forums to extract the account information from 5147 unique opioid suppliers. By the time Agora was shut down in August 2015, 916 opioid suppliers were found, with an average number of listings of 13 per supplier. We observed that the opioid suppliers with most listings is *mikesales*, which contributes to 817 opioid listings for the marketplace Darkbay, whereas the opioid supplier ID which was observed in most marketplaces is *DeepMeds*, which posted the similar listings of buprenorphine, codeine, and narcotic in 6 different marketplaces from 2014 to 2020. The average number of listings that a legit supplier posts on Amazon is approximately 37 [34,35], which is far less than those posted by suppliers in anonymous marketplaces. It is possible that those suppliers in darknet marketplaces are hidden under an anonymous environment with little to no limitations.

To better understand the potential bundling relationship of opioid suppliers across different marketplaces, we calculated the Jaccard similarity coefficient between the suppliers in different marketplaces (Figure 6). We found that 182 opioid supplier IDs appeared in both the marketplaces Evolution and Agora from January 2014 to July 2015. In particular, we observed that 84 opioid supplier IDs synchronized similar product listings in both marketplaces at the same time. This might be because the suppliers tend to promote their products across various marketplaces and to increase sales.

Figure 6. Co-occurrence of the same opioid suppliers across different anonymous marketplaces.



Inspired by the work [36] investigating the supplier migration phenomenon between underground marketplaces, we also evaluated the migrant suppliers who, for the first time, began to trade in a new market, *m'*, after the closure of marketplace *m*. To this end, we first collected the marketplace's lifespan using the Gwern archive [37], as shown in Table 1, and then

compared the supplier lists in each marketplace to investigate supplier migration. We found that 28 and 35 suppliers in Evolution moved to Agora and Alphabay, respectively, after March 2015 when Evolution was shut down, which is aligned with the finding of users' migration between these 3 markets by El Bahrawy et al [36], who used a bitcoin address instead of supplier IDs for user matching. In addition, we observed that 10 and 9 suppliers in Apollon and Empire, respectively, migrated from Agora 3 years after it shut down. Some of those suppliers' IDs (ie, *A1CRACK* and *DiazNL*) are neither common words nor have special meaning. We hypothesized that these IDs might be linked to the same supplier. Those suppliers kept using the same IDs for years to gain reputation and familiarity from buyers.

In addition, 204 suppliers were reported as scammers in the anonymous forums of the Evolution, SilkRoad, Pandora, and The Hub. It is not surprising to find that the top 3 marketplaces and forums that found the most scam reports were Evolution, Silk Road, and Pandora, as the scam reports mostly come from the associated forums of the marketplaces [38].

Characteristics of the Drug Transaction

When inspecting the advertised origins and the acceptable shipping destinations on the opioid listings from 7 marketplaces (Apollon, Avaris, Alphabay, Hydra, Pandora, Empire, and Versus), we observed that most of the opioid commodities were shipped from the United States, followed by the United Kingdom, Germany, Netherlands, and Canada (Table 7). This finding is roughly in line with the observation of 57 opioid vendors' origin in a marketplace named Cryptomarket during the period of October 2015 through April 2016, which was reported by Duxbury et al [6]. In addition, we did not observe big changes in the top opioid commodity origins from 2014 to 2020. Particularly, as shown in Table 8, the United States and the United Kingdom are always in the top 5 advertised origins among years.

Table 7. The advertised origin countries.

Name of country	Percentage of origin countries in opioid listings, n (%)
United States	3520 (35.3)
United Kingdom	1648 (17.1)
Germany	1459 (14.7)
Netherlands	897 (9)
Canada	622 (6.2)
France	479 (4.8)
Australia	398 (4)
India	200 (2)
Spain	160 (1.6)
Sweden	80 (0.8)
Japan	73 (0.7)
Italy	60 (0.6)
Singapore	55 (0.6)
Belgium	51 (0.5)
Switzerland	50 (0.5)
Portugal	22 (0.2)
Afghanistan	18 (0.2)
Denmark	17 (0.2)
Czech Republic	14 (0.1)
China	11 (0.1)

Table 8. Top 5 advertised origin countries according to different years. Note that data for 2020 only included data from January to March.

2014		2015		2019		2020	
Country	Number of appearance in listings	Country	Number of appearance in listings	Country	Number of appearance in listings	Country	Number of appearance in listings
United States	778	United States	603	United States	1394	United States	680
Germany	646	France	188	United Kingdom	1269	Netherlands	250
Netherlands	145	Canada	151	Germany	537	United Kingdom	185
United Kingdom	142	Australia	112	Netherlands	451	Germany	176
Canada	131	United Kingdom	96	Canada	287	Australia	75

Considering the shipping destination, we observed that the majority of opioid commodities were shipped worldwide 36.37% (5654/15,546), followed by shipping to the United States only 19.35% (3008/15,546), Europe only 10.52% (1635/15,546), and the United Kingdom only 5.46% (849/15,546).

To understand customer satisfaction, we conducted a sentiment analysis on 624 review posts related to 190 opioid suppliers from 4 marketplaces: Agora, Alphabay, Pandora, and Evolution. We observed that 145 opioid suppliers had 378 positive reviews, whereas 102 opioid suppliers had at least one negative review. For instance, the opioid supplier from the SilkRoad with the user ID *c63amg* received a satisfaction rating of 76%, even though they had the most negative reviews (n=11). We notice that their negative reviews mostly came from one buyer in late 2012 and early 2013, who complained “his Heroin is getting from order to order worse.”

As observed in our data set, the opioid suppliers in the marketplaces Evolution, Pandora, and Silk Road accepted escrow as a method of payment. However, most of the suppliers only used escrow for small orders. This shows the weak platform trust of opioid suppliers. In fact, the shutdown of the marketplace Evolution was discovered to be an exit scam, with the site’s operators shutting down abruptly to steal the approximately US $12 million in bitcoins that it was holding as an escrow [39].

Discussion

Principal Findings

Our study identified 41,000 opioid trade–related marketplace listings and forum posts by analyzing more than 1 million listings and posts in multiple anonymous marketplaces and forums, which is the largest underground opioid trading data set ever reported. We found evidence through extensive analyses of the anonymous online market of pervasive supply, which fuels the international opioid epidemic. Nontraditional methods, as presented here by studying the online supply chain, present a novel approach for governmental and other large-scale solutions. When interpreted by professionals, our initial results demonstrate useful findings and may be used downstream by law enforcement and public policy makers for impactful structural interventions to the opioid crisis. Although a large body of current research is focused on pathways for treatment of opioid use disorder and analyzing deaths per treatment capacity of substance use providers, these research areas are limited to the demand side of the opioid epidemic [40,41]. We believe that the findings and pattern analyses presented here, which place concentration on the supply side, might suggest a new direction to focus and will serve as a useful complement to current research conducted within the domain of addiction medicine.

Limitations

We acknowledge some limitations of our study. For example, there might be varying types of heroin or fentanyl, but we could not subcategorize them due to the lack of precise ontology. Addressing this challenge requires deep domain knowledge and expertise, which is constantly evolving. Another limitation is pointed out in the paper that multiple online suppliers might belong to the same vendor. This problem might be addressed by studying the product overlapping patterns over time to merge suppliers, which might reveal more interesting hierarchical clustering patterns. Another important source of information is the trading cash flow, which is recorded in the block chain and might contribute to a comprehensive view of the supply-demand relationship. We did not include such analyses due to the time and scope constraints, and it is a topic that we plan to investigate further.

Conclusions

In our study, a total of 248,359 listings from 10 anonymous online marketplaces and 1,138,961 traces (ie, threads of posts) from 6 underground forums were collected. Among them, we identified 28,106 opioid product listings and 13,508 opioid-related promotional and review forum traces from 5147 unique opioid suppliers’ IDs and 2778 unique opioid buyers’ IDs. Our study characterized opioid suppliers (eg, activeness and cross-market activities), commodities (eg, popular items and their evolution), and transactions (eg, origins and shipping destination) in anonymous marketplaces and forums, which enabled a greater understanding of the underground trading activities involved in international opioid supply and demand.

To the best of our knowledge, a comprehensive overview of the opioid supply chain in the anonymous online marketplaces and forums, as well as a measurement study of trading activities, is still an open research challenge. This is the first study to measure and characterize opioid trading in anonymous online marketplaces and forums. From our measurement, we concluded that anonymous online marketplaces and forums provided easy-access platforms for global opioid supply. These findings characterizing mass opioid suppliers, commodities, and transactions on anonymous marketplaces and forums can enable law enforcement, policy makers, and invested health care stakeholders to better understand the scope of opioid trading activities and provide insight into this new type of opioid supply chain.

Conflicts of Interest

None declared.

References

1. Rudd RA, Aleshire N, Zibbell JE, Gladden RM. Increases in drug and opioid overdose deaths--United States, 2000-2014. MMWR Morb Mortal Wkly Rep 2016 Jan 1;64(50-51):1378-1382 [FREE Full text] [doi: 10.15585/mmwr.mm6450a3] [Medline: 26720857]
2. Drug Overdose Data. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/products/vsrr/drug-overdose-data.htm [accessed 2020-03-15]
3. Barratt M. Silk Road: eBay for drugs. Addiction 2012 Mar;107(3):683. [doi: 10.1111/j.1360-0443.2011.03709.x] [Medline: 22313109]
4. Pacula RL, Powell D. A supply-side perspective on the opioid crisis. J Pol Anal Manage 2018 Feb 16;37(2):438-446. [doi: 10.1002/pam.22049]
5. Christin N. Traveling the Silk Road: a Measurement Analysis of a Large Anonymous Online Marketplace. In: 22nd international conference on World Wide Web. 2013 May 16 Presented at: WWW'13; January 23-25, 2013; Rio de Janeiro, Brazil URL: https://doi.org/10.1145/2488388.2488408 [doi: 10.1145/2488388.2488408]
6. Duxbury SW, Haynie DL. The network structure of opioid distribution on a darknet cryptomarket. J Quant Criminol 2017 Jun 8;34(4):921-941. [doi: 10.1007/s10940-017-9359-4]
7. Lee S, Yoon C, Kang H, Kim Y, Kim Y, Han D, et al. Cybercriminal Minds: an Investigative Study of Cryptocurrency Abuses in the Dark Web. In: Network and Distributed Systems Security (NDSS) Symposium. 2019 Presented at: NDSS'19; May 6-7, 2019; San Diego, USA. [doi: 10.14722/ndss.2019.23055]
8. Soska K, Christin N. Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In: Proceedings of the 24th USENIX Security Symposium (USENIX Security 2015). 2004 Aug 12 Presented at: SS'04; June 7-9, 2015; Washington DC p. 33-48 URL: https://dl.acm.org/doi/abs/10.5555/2831143.2831146
9. Zhang Y, Fan Y, Song W, Hou S, Ye Y, Li X, et al. Your Style Your Identity: Leveraging Writing and Photography Styles for Drug Trafficker Identification in Darknet Markets Over Attributed Heterogeneous Information Network. In: The Web Conference. 2019 13 Presented at: WC'19; July 23-27, 2019; San Francisco. [doi: 10.1145/3308558.3313537]
10. Van Wegberg R, Tajalizadehkhoob S, Soska K, Akyazi U, Ganan CH, Klievink B, et al. Plug and prey? Measuring the commoditization of cybercrime via online anonymous markets. In: 27th USENIX security symposium (USENIX security 18). 2018 Presented at: USENIX Association; USENIX Association, Baltimore, USA p. 1009-1026 URL: https://dl.acm.org/doi/10.5555/3277203.3277279
11. Balsamo D, Bajardi P, Salomone A, Schifanella R. Patterns of routes of administration and drug tampering for nonmedical opioid consumption: data mining and content analysis of reddit discussions. J Med Internet Res 2021 Jan 4;23(1):e21212 [FREE Full text] [doi: 10.2196/21212] [Medline: 33393910]
12. Cherian R, Westbrook M, Ramo D, Sarkar U. Representations of codeine misuse on instagram: content analysis. JMIR Public Health Surveill 2018 Mar 20;4(1):e22 [FREE Full text] [doi: 10.2196/publichealth.8144] [Medline: 29559422]
13. Mackey T, Kalyanam J, Klugman J, Kuzmenko E, Gupta R. Solution to detect, classify, and report illicit online marketing and sales of controlled substances via twitter: using machine learning and web forensics to combat digital opioid access. J Med Internet Res 2018 Apr 27;20(4):e10029 [FREE Full text] [doi: 10.2196/10029] [Medline: 29613851]
14. Nasralah T, El-Gayar O, Wang Y. Social media text mining framework for drug abuse: development and validation study with an opioid crisis case analysis. J Med Internet Res 2020 Aug 13;22(8):e18350 [FREE Full text] [doi: 10.2196/18350] [Medline: 32788147]
15. Gilbert M, Dasgupta N. Silicon to syringe: cryptomarkets and disruptive innovation in opioid supply chains. Int J Drug Policy 2017 Aug;46:160-167. [doi: 10.1016/j.drugpo.2017.05.052] [Medline: 28735772]
16. Quintana P, Ventura M, Grifell M, Palma A, Galindo L, Fornís I, et al. The hidden web and the fentanyl problem: detection of ocfentanil as an adulterant in heroin. Int J Drug Policy 2017 Feb;40:78-83. [doi: 10.1016/j.drugpo.2016.10.006] [Medline: 27889114]
17. Caudevilla F, Ventura M, Fornís I, Barratt MJ, Vidal C, Lladanosa CG, et al. Results of an international drug testing service for cryptomarket users. Int J Drug Policy 2016 Sep;35:38-41. [doi: 10.1016/j.drugpo.2016.04.017] [Medline: 27239011]

18. Dasgupta N, Freifeld C, Brownstein JS, Menone CM, Surratt HL, Poppish L, et al. Crowdsourcing black market prices for prescription opioids. J Med Internet Res 2013;15(8):e178 [FREE Full text] [doi: 10.2196/jmir.2810] [Medline: 23956042]
19. Yuan K, Lu H, Liao X, Wang X. Reading Thieves' Cant: Automatically Identifying and Understanding Dark Jargons from Cybercrime Marketplaces. In: 27th USENIX security symposium. 2018 Aug 15 Presented at: 27th USENIX security symposium (USENIX security 18); USENIX Association Baltimore, USA p. 1027-1041 URL: https://dl.acm.org/doi/10.5555/3277203.3277280
20. Gwern B, Nicolas C, David D, Rasmus MA, StExo, El P, Anonymous, Sohhlz, Whom, et al. Dark Net Market archives, 2011-2015. GWERN. URL: https://www.gwern.net/DNM-archives [accessed 2020-03-10]
21. Pergolizzi JV, LeQuang JA, Taylor R, Raffa RB, NEMA Research Group. The 'Darknet': the new street for street drugs. J Clin Pharm Ther 2017 Dec 15;42(6):790-792. [doi: 10.1111/jcpt.12628] [Medline: 28921578]
22. Ladegaard I. Crime displacement in digital drug markets. Int J Drug Policy 2019 Jan;63:113-121. [doi: 10.1016/j.drugpo.2018.09.013] [Medline: 30572247]
23. Dozens Of Online Dark Markets Seized Pursuant To Forfeiture Complaint Filed In Manhattan Federal Court In Conjunction With The Arrest Of The Operator Of Silk Road 2. The Department of Justice, Southern District of New York. URL: https://www.justice.gov/usao-sdny/pr/dozens-online-dark-markets-seized-pursuant-forfeiture-complaint-filed-manhattan-federal/ [accessed 2020-03-10]
24. Packaging WG. Python language bindings for Selenium WebDriver. Selenium 3. URL: https://pypi.org/project/selenium/ [accessed 2020-03-10]
25. DEA Intelligence Report: Drug Slang Code Words (2017). Drug Enforcement Administration (DEA). URL: https://cesar.umd.edu/sites/cesar.umd.edu/files/pubs/dea-drug-slang-code-words-may2017.pdf [accessed 2020-03-10]
26. Simpson SS, Adams N, Brugman CM, Conners TJ. Detecting novel and emerging drug terms using natural language processing: a social media corpus study. JMIR Public Health Surveill 2018 Jan 8;4(1):e2 [FREE Full text] [doi: 10.2196/publichealth.7726] [Medline: 29311050]
27. Zittleman L, Curcija K, Sutter C, Dickinson LM, Thomas J, de la Cerda D, et al. Building capacity for medication assisted treatment in rural primary care practices: the IT matttrs practice team training. J Prim Care Community Health 2020;11:2150132720953723 [FREE Full text] [doi: 10.1177/2150132720953723] [Medline: 32909491]
28. McAuley J. Amazon Product Data. Amazon Review. URL: http://jmcauley.ucsd.edu/data/amazon/ [accessed 2020-03-10]
29. LC-John. Yahoo-Answers-Topic-Classification-Dataset. GitHub. 2018. URL: https://github.com/LC-John/Yahoo-Answers-Topic-Classification-Dataset [accessed 2020-03-10]
30. MAchine Learning for LanguagE Toolkit (MALLET). UMass Amherst. 2002. URL: http://mallet.cs.umass.edu/index.php [accessed 2020-03-10]
31. Portnoff R, Afroz S, Durrett G. Tools for Automated Analysis of Cybercriminal Markets. In: the 26th International Conference on World Wide Web. 2017 Presented at: WWW'17; July 3-5, 2017; Perth, Australia. [doi: 10.1145/3038912.3052600]
32. Sentiment Polarity Analysis. GitHub. URL: https://github.com/chaoming0625/ [accessed 2020-03-10]
33. Heroin and cocaine prices in Europe and USA. United Nations Office on Drugs and Crime. URL: https://dataunodc.un.org/drugs/heroin_and_cocaine_prices_in_eu_and_usa-2017 [accessed 2020-03-10]
34. Number of Sellers on Amazon Marketplace. Marketplace Pluse. URL: https://www.marketplacepulse.com/amazon/number-of-sellers [accessed 2020-03-10]
35. Emily D. Amazon Statistics You Should Know. Big Commerce. URL: https://www.bigcommerce.com/blog/amazon-statistics/#10-fascinating-amazon-statistics-sellers-need-to-know-in-2020 [accessed 2020-03-10]
36. ElBahrawy A, Alessandretti L, Rusnac L, Goldsmith D, Teytelboym A, Baronchelli A. Collective dynamics of dark web marketplaces. Sci Rep 2020 Nov 2;10(1):18827 [FREE Full text] [doi: 10.1038/s41598-020-74416-y] [Medline: 33139743]
37. Darknet Market Mortality Risks. Gwern. URL: https://www.gwern.net/DNM-survival [accessed 2020-03-10]
38. Cox J. Reputation Is Everything: the Role of Ratings, Feedback and Reviews in Cryptomarkets. The Internet and Drug Markets. European Monitoring Centre for Drugs and Drug Addiction: Insights 21; 2016:49-54 URL: https://www.emcdda.europa.eu/publications/insights/internet-drug-markets_en
39. Greenberg A. The Dark Web's Top Drug Market, Evolution, Just Vanished. Wired Magazine. 2015. URL: https://www.wired.com/2015/03/evolution-disappeared-bitcoin-scam-dark-web/ [accessed 2020-03-10]
40. Champagne-Langabeer T, Swank MW, Langabeer JR. Routes of non-traditional entry into buprenorphine treatment programs. Subst Abuse Treat Prev Policy 2020 Jan 20;15(1):6-10 [FREE Full text] [doi: 10.1186/s13011-020-0252-z] [Medline: 31959194]
41. Langabeer JR, Gourishankar A, Chambers KA, Giri S, Madu R, Champagne-Langabeer T. Disparities between US opioid overdose deaths and treatment capacity: a geospatial and descriptive analysis. J Addict Med 2019;13(6):476-482. [doi: 10.1097/ADM.0000000000000523] [Medline: 30844879]

Abbreviations

API: Application Programming Interface
MALLET: Machine Learning for Language Toolkit

Edited by G Eysenbach; submitted 21.09.20; peer-reviewed by E Yom-Tov, P Bajardi, M Torii; comments to author 23.11.20; revised version received 15.01.21; accepted 16.01.21; published 17.02.21.

Please cite as:
Li Z, Du X, Liao X, Jiang X, Champagne-Langabeer T
Demystifying the Dark Web Opioid Trade: Content Analysis on Anonymous Market Listings and Forum Posts
J Med Internet Res 2021;23(2):e24486
URL: http://www.jmir.org/2021/2/e24486/
doi:10.2196/24486
PMID:33595442

Original Paper

A Virtual Ward Model of Care for Patients With COVID-19: Retrospective Single-Center Clinical Study

Olivia R Ferry[1*], MD; Emma C Moloney[1*], MBBS; Owen T Spratt[1], MD; Gerald F M Whiting[1], FRCPI; Cameron J Bennett[1], FRACP

Metro North Hospital and Health Service, Brisbane, Australia

[*]these authors contributed equally

Corresponding Author:
Cameron J Bennett, FRACP
Metro North Hospital and Health Service
Royal Brisbane and Women's Hospital, Butterfield Street, Herston
Brisbane, 4006
Australia
Phone: 61 36468111
Email: cameron.bennett@health.qld.gov.au

Abstract

Background: COVID-19 has necessitated the implementation of innovative health care models in preparation for an influx of patients. A virtual ward model delivers clinical care remotely to patients in isolation. We report on an Australian cohort of patients with COVID-19 treated in a virtual ward.

Objective: The aim of this study was to describe and evaluate the safety and efficacy of a virtual ward model of care for an Australian cohort of patients with COVID-19.

Methods: Retrospective clinical assessment was performed for 223 patients with confirmed COVID-19 treated in a virtual ward in Brisbane, Australia, from March 25 to May 15, 2020. Statistical analysis was performed for variables associated with the length of stay and hospitalization.

Results: Of 223 patients, 205 (92%) recovered without the need for escalation to hospital care. The median length of stay in the virtual ward was 8 days (range 1-44 days). In total, 18 (8%) patients were referred to hospital, of which 6 (33.3%) were discharged after assessment at the emergency department. Furthermore, 12 (5.4%) patients were admitted to hospital, of which 4 (33.3%) required supplemental oxygen and 2 (16.7%) required mechanical ventilation. No deaths were recorded. Factors associated with escalation to hospital care were the following: hypertension (odds ratio [OR] 3.6, 95% CI 1.28-9.87; *P*=.01), sputum production (OR 5.2, 95% CI 1.74-15.49; *P*=.001), and arthralgia (OR 3.8, 95% CI 1.21-11.71; *P*=.02) at illness onset and a polymerase chain reaction cycle threshold of ≤20 on a diagnostic nasopharyngeal swab (OR 5.0, 95% CI 1.25-19.63; *P*=.02).

Conclusions: Our results suggest that a virtual ward model of care to treat patients with COVID-19 is safe and efficacious, and only a small number of patients would potentially require escalation to hospital care. Further studies are required to validate this model of care.

(J Med Internet Res 2021;23(2):e25518) doi:10.2196/25518

KEYWORDS

COVID-19; efficacy; hospital; innovation; model; remote care; safety; telemedicine; virtual health care; virtual ward

Introduction

On March 11, 2020, the World Health Organization declared COVID-19, a respiratory infection due to SARS-CoV-2, as a global pandemic [1]. A key consideration in this pandemic has been the management of the rapid influx of patients with COVID-19. The subsequent strain on health care systems has acted as a catalyst for increasing the implementation of telemedicine [2]. Telemedicine refers to health care provision through information technologies and telecommunication systems [3]. A Cochrane review [4] concluded that telemedicine can have equivalent outcomes to those of in-person care. However, the implementation of novel telemedicine approaches can be challenging, since adaptation of both staff and patients is required. During COVID-19, telemedicine has been used to triage, treat, and coordinate care provision to patients with

COVID-19 to improve health care access, reduce disease transmission, and optimize resource allocation [5-7]. A virtual ward delivers hospital-level care to patients in the community through telemedicine. Through the provision of timely multidisciplinary care, virtual ward models have reduced emergency department presentations and hospital admissions [8,9]. These outcomes are desirable in a pandemic, where the judicious use of limited health care resources is critical.

To provide care to patients safely and effectively through a virtual care model, it is important to understand the clinical course of COVID-19 [2]. Several meta-analyses of published cohort studies have described the most common initial symptoms of COVID-19, including cough and fatigue [10-12]. Common comorbidities identified in patients with confirmed COVID-19 are hypertension (15.6%), diabetes (7.7%), and cardiovascular disease (4.7%) [11]. The most common laboratory abnormalities include an increased C-reactive protein level (68.6%), lymphopenia (57.4%), and an increased lactate dehydrogenase level (51.6%) [12]. The reported clinical spectrum of COVID-19 is broad, ranging from asymptomatic infection and mild upper respiratory tract illnesses to severe pneumonia and critical multiorgan failure [13]. The current literature suggests that approximately 80% of cases are mild [13]. However, out of 44,500 cases of COVID-19 in China, 14% of patients experienced severe disease with hypoxia and 5% of critical cases experienced respiratory failure, shock, or multiorgan dysfunction [13]. Mortality rates vary by region and the data collection method. Initial studies in China have reported mortality rates of 2.3%-3.6%, with a higher mortality associated with a higher age or the presence of comorbidities [11-13]. This emphasizes the potential for most patients with COVID-19 to be treated in lower acuity settings with monitoring for disease exacerbation.

Predictors of disease exacerbation during acute COVID-19 have been proposed in early retrospective cohort studies on patients with COVID-19 pneumonia or severe disease [14-18]. Baseline characteristics such as increasing age, male sex, and comorbidities confer a greater risk of severe disease and mortality [14,17,19,20]. In particular, chronic lung disease, cardiovascular disease, hypertension, diabetes mellitus, and immunosuppression have been proposed as risk factors [21-23]. In severe disease, a higher incidence of dyspnea (approximately 67%) has been reported in those requiring admission to the intensive care unit (ICU) [14,18]. Additionally, new-onset dyspnea may reflect the development of COVID-19 pneumonia. In cohort studies on COVID-19 pneumonia, the median time to dyspnea onset has been reported as 5-8 days after initial symptom onset [15,16]. Additionally, high body temperatures (≤39°C) are associated with an increased likelihood of acute respiratory distress syndrome (ARDS) [17]. The time to deterioration is notable with a median of 8-12 days from illness onset to ARDS and 10 days to ICU admission [14-16,19]. While further data are needed, this second week of acute illness likely represents a high-risk period for disease exacerbation, which may bolster clinical decision making regarding hospitalization.

The aims of this study were as follows: (1) to describe the clinical characteristics of an Australian cohort of patients with COVID-19, (2) to evaluate the clinical care provided to this cohort through a virtual ward model, and (3) to identify any possible predictors of deterioration.

Methods

Study Design

A retrospective single-center clinical assessment was performed for patients admitted to the Metro North Virtual Ward from March 25 to May 15, 2020. This study was deemed at low/negligible risk by the Royal Brisbane and Women's Hospital Human Research Ethics Committee. No formal power calculations were performed, owing to the inclusion of all patients meeting the study criteria.

Study Population

All patients admitted to the virtual ward during the specified period were assessed in accordance with the following inclusion criterion: a laboratory-confirmed diagnosis of COVID-19 through polymerase chain reaction (PCR) detection of SARS-CoV-2 RNA on a diagnostic nasopharyngeal swab (NPS). Patients were excluded if they had a preliminary positive or inconclusive PCR result but a negative result on subsequent confirmatory PCR testing. Patients were admitted to the virtual ward from the community after notification of a positive PCR result by the Metro North Public Health Unit, Herston, Australia, or following hospital discharge, in cases of confirmed disease.

Virtual Care

Patients remained in out-of-hospital isolation during their virtual ward admission with nursing observations obtained through telephonic consultations. Virtual ward staff were located in a secure dedicated hospital workspace with medical records maintained in accordance with local hospital procedures and protocols. Patients were risk-stratified by age, comorbidities, and symptom burden to determine the frequency of telephonic consultations: low-risk patients, once daily; high-risk patients, twice daily. Observations were structured to monitor patient symptoms and identify potential deterioration. During each consultation, patients were asked to rate (on a scale of none, mild, moderate, or severe) the following symptoms: shortness of breath, cough, fatigue, sputum production, nausea/vomiting, headache, myalgia, and sore throat. These symptoms were numerically scored at each review. Patients' general well-being, social situation, and adherence to isolation were also assessed.

Clinical reviews were conducted by medical officers when the following prespecified escalation criteria were met: (1) the patient reported severe symptoms related to shortness of breath, cough, or fatigue; (2) symptoms became more severe either on 1 observation of patients aged >65 years and having comorbidities or over 2 observations in those without comorbidities; or (3) any staff or patient concerns regarding disease exacerbation. If required, hospital referral was arranged for further assessment. All patients were reviewed by a medical officer prior to discharge. Multidisciplinary care was provided, with pharmacists ensuring patient access to medications and social workers offering psychosocial support.

In accordance with the Communicable Diseases Network Australia (CDNA) COVID-19 guidelines [24], patients were

discharged and released from self-isolation on meeting the following recovery criteria: (1) 10 days since symptom onset and resolution of all symptoms of acute illness in the past 72 hours and (2) 10 days since hospital discharge and resolution of all symptoms of acute illness in the past 72 hours, if hospitalized for severe COVID-19.

Data Collection

Data on patient demographics, epidemiological history, comorbidities, medication history, COVID-19 symptoms, clinical reviews, pathology results, hospital assessment, and treatment outcomes were collected from existing medical records.

Data Analysis

We expressed descriptive statistics as number (%) values for categorical data and median or mean (range) values for continuous variables. We performed Pearson χ^2 tests to explore the risk factors among patients requiring hospital referral and those with a virtual ward stay >7 days. We used Fisher exact tests when event counts were <5. Missing data were not imputed in the analyses. Furthermore, we calculated odds ratio (OR) and corresponding 95% CI values. All tests were two-sided, with a *P* value <.05 considered significant. Data were not adjusted for multiple testing; hence, findings should be considered to be descriptive and should not be used to infer definitive effects. SPSS (version 26.0, IBM Corp) was used for the analyses.

Results

Patient Characteristics

A total of 223 patients with a median age of 45 (range 14-78) years (female n=118, 52.9%) were assessed in this study (Table 1). This included 2 patients aged <18 years. Almost half (n=100, 44.8%) of the patients had a comorbidity, with hypertension (n=38, 17%) and asthma (n=24, 10.8%) being the most common manifestations. A total of 178 (79.8%) cases were epidemiologically linked to overseas travel (Figure 1), the most common destinations being the United Kingdom (n=68, 38%) and the United States (n=30, 17%). Furthermore, 16 (7.2%) patients had traveled on cruise ships. The most common COVID-19 symptoms upon presentation were cough (n=163, 73.1%), fever (n=117, 52.5%), and headache (n=103, 46.2%). Prior to virtual ward admission, 100 (44.8%) patients were assessed by a medical practitioner, either in person or through telemedicine, 21 (9.4%) had undergone chest radiography, and 22 (9.9%) had received laboratory blood tests. Initial diagnostic PCR detected SARS-CoV-2 RNA at a median cycle threshold (Ct) of 23.88.

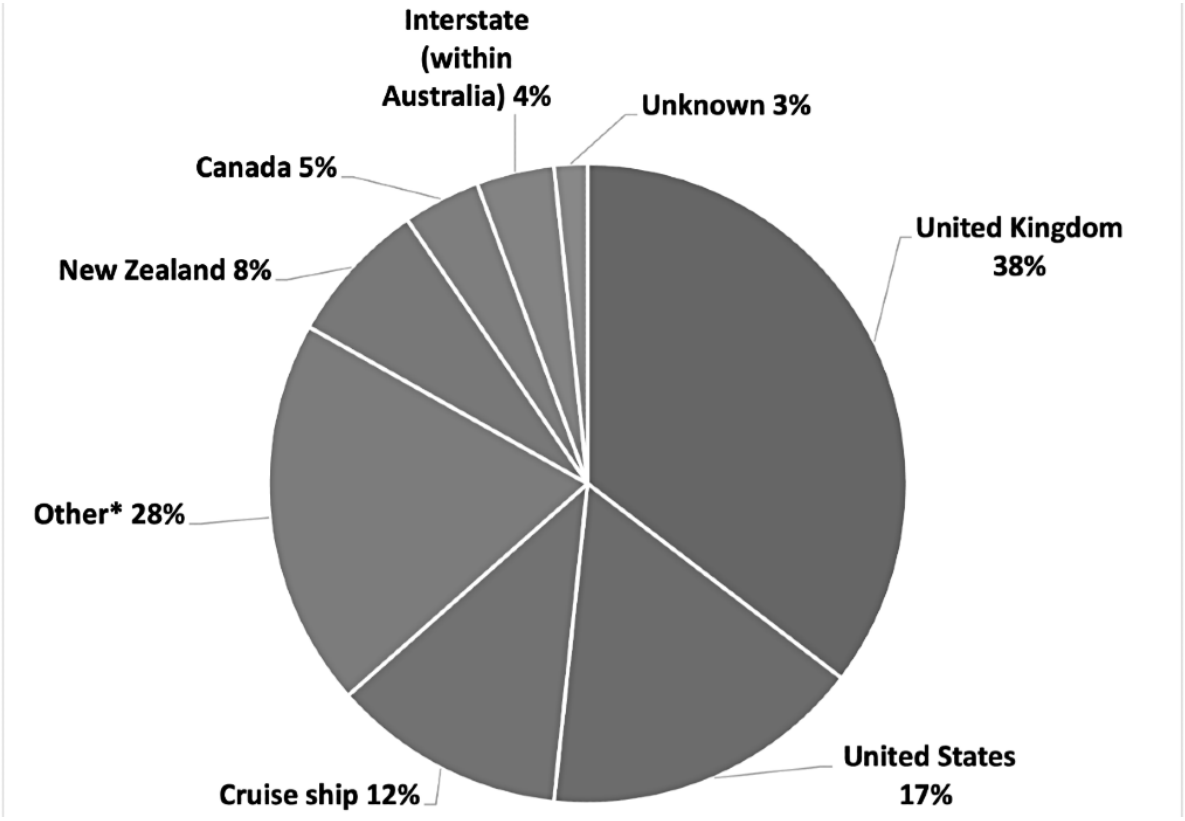
Table 1. Baseline characteristics of the study population (N=223).

Characteristics	All patients	Patients not referred to hospital (n=205)	Patients referred to hospital (n=18)
Median age, years (range)	45.0 (14-78)	42.0 (14-78)	54.0 (23-71)
Female sex, n (%)	118 (52.9)	108 (52.7)	10 (55.6)
High risk[a], n (%)	63 (28.3)	54 (26.3)	9 (50)
Transmission source, n (%)			
Overseas travel	178 (79.8)	167 (81.5)	11 (61.1)
Contact with a confirmed case	47 (21.1)	43 (21)	4 (22.2)
Locally acquired	43 (19.3)	35 (17.1)	8 (44.4)
Unknown	3 (1.3)	3 (1.5)	0 (0)
Comorbidities, n (%)			
Any	100 (44.8)	87 (42.4)	13 (72.2)
Hypertension	38 (17)	31 (15.1)	7 (38.9)
Asthma	24 (10.8)	23 (11.2)	1 (5.6)
Diabetes mellitus	13 (5.8)	12 (5.9)	1 (5.6)
Immunosuppression[b]	6 (2.7)	5 (2.4)	1 (5.6)
Chronic obstructive pulmonary disease	3 (1.3)	3 (1.5)	0 (0)
Medication (angiotensin-converting enzyme inhibitors angiotensin II receptor blockers, total 135), n (%)	25 (11.2)	19 (9.3)	6 (33.3)
Initial symptoms at onset, n (%)			
Cough	163 (73.1)	149 (72.7)	14 (77.8)
Fever	117 (52.5)	104 (50.7)	13 (72.2)
Headache	103 (46.2)	97 (47.3)	6 (33.3)
Sore throat	97 (43.5)	90 (43.9)	7 (38.9)
Fatigue	84 (37.7)	74 (36.1)	10 (55.6)
Rhinorrhea	82 (36.8)	76 (37.1)	6 (33.3)
Myalgia	78 (35)	70 (34.1)	8 (44.4)
Shortness of breath	52 (23.3)	46 (22.4)	6 (33.3)
Nausea/vomiting	37 (16.6)	33 (16.1)	4 (22.2)
Diarrhea	37 (16.6)	35 (17.1)	2 (11.1)
Anosmia	36 (16.1)	35 (17.1)	1 (5.6)
Ageusia	30 (13.5)	27 (13.2)	3 (16.7)
Sputum	24 (10.8)	18 (8.8)	6 (33.3)
Arthralgia	24 (10.8)	19 (9.3)	5 (27.8)
Chest tightness	20 (9)	17 (8.3)	3 (16.7)
Initial presentation			
Initial assessment by a medical practitioner, n (%)	100 (44.8)	90 (43.9)	10 (55.6)
Initially admitted to hospital prior to virtual ward admission, n (%)	32 (14.3)	29 (14.1)	3 (16.7)
Median time from symptom onset to initial nasopharyngeal swab (n=177), days (range)	4 (0-23)	4 (0-23)	1 (0-5)
Median cycle threshold of polymerase chain reaction analysis of initial nasopharyngeal swabs (n=135) (range)	23.88 (11-36)	24.00 (11-36)	18.04 (14.6-33)
Chest radiography, n (%)	21 (9.4)	20 (9.8)	1 (5.6)
Blood tests, n (%)	22 (9.9)	21 (10.3)	2 (11.1)

[a]High risk was defined as age 65-85 years with any comorbidity or age 49-65 years with chronic lung disease, cardiovascular disease, immunosuppression, diabetes, or hypertension.

[b]Immunosuppression was defined as patients taking immunosuppressive medication or having a primary immunodeficiency.

Figure 1. Cases epidemiologically linked to overseas or interstate travel (n=178). *Others: Argentina, Cuba, Egypt, Hong Kong, Indonesia, Japan, Myanmar, Norway, the Philippines, Singapore, South Africa, Sweden, Turkey, and the United Arab Emirates.



Virtual Ward Outcomes

Of the 223 patients in the virtual ward, 205 (92%) were discharged after clinical recovery without escalation to hospital care (Table 2). The median virtual ward admission length was 8 days (range 1-44 days). The median time to clinical recovery was 16 days (range 10-52 days). A total of 18 (8.1%) patients were referred for in-person hospital assessment (Table 3), of which 6 (33.3%) were assessed at the emergency department and discharged back to the virtual ward after review. The remaining 12 (66.7%) patients were admitted to an in-patient ward, of which 2 (16.7%) were admitted to the ICU and required mechanical ventilation (Table 3). The average length of hospital stay was 3.5 days (range 1-15 days) when admission to the ICU was not required. We recorded no mortality upon discharging the 223 patients assessed in this study.

Several factors were associated with a length of virtual ward stay >7 days. These included having any comorbidity (OR 2.0, 95% CI 1.15-3.40; *P*=.01), being classified as high risk on admission (OR 2.2, 95% CI 1.16-4.00; *P*=.02), or being hospitalized prior to virtual ward admission (OR 2.6, 95% CI 1.10-5.99; *P*=.03). Initial COVID-19 symptoms of cough (OR 2.2, 95% CI 1.22-4.10; *P*=.008), fevers or night sweats (OR 2.2, 95% CI 1.26-3.70; *P*=.005), and diarrhea (OR 2.3, 95% CI 1.06-5.07; *P*=.03) were also associated with a length of virtual ward stay >7 days.

Table 2. Virtual ward patient outcomes (N=223).

Outcome	Patients
Median length of stay, days (range)	8.0 (1-44)
Median time to clinical recovery[a], days (range)	16 (10-52)
Discharged without complication, n (%)	205 (91.9)
Requiring hospital assessment, n (%)	18 (8.1)
Admitted to the in-patient ward, n (%)	12 (5.4)
Mean length of in-patient hospitalization, if admission to the intensive care unit was not required[a], days (range)	3.5 (1-15)
Admitted to the intensive care unit, n (%)	2 (0.9)
Mortality, n (%)	0 (0)

[a]Time from symptom onset to clinical recovery in accordance with the Communicable Diseases Network Australia guidelines (at least 10 days since symptom onset and 72 hours of being asymptomatic).

Table 3. Clinical characteristics of patients requiring hospital care upon admission to the virtual ward (N=18).

Characteristics	Hospitalized patients
Median age, years (range)	54 (23-71)
Female sex, n (%)	10 (55.6)
High risk[a], n (%)	9 (50)
Median day of illness upon referral to hospital, days (range)	8.50 (3-20)
Previous medical review on initial presentation, n (%)	10 (55.6)
Initial median polymerase chain reaction cycle threshold on diagnostic nasopharyngeal swab tests, n (range)	18.04 (14.61-33)
Primary reason for referral, n (%)	
Shortness of breath	10 (55.6)
New or ongoing fevers	4 (22.2)
Chest pain or chest tightness	3 (16.7)
Hospital assessment on presentation	
Median oxygen saturation (n=13), % of ambient air (range)	96 (88-100)
Median heart rate (n=10), beats per minute (range)	80 (57-105)
Median respiratory rate (n=9), breaths per minute (range)	19 (16-28)
Fever (>37.5°C; n=11), n (%)	4 (22.2)
Chest auscultation findings (n=15), n (%)	
Clear	9 (60)
Unilateral crackles	2 (13.3)
Bilateral crackles	4 (26.7)
Chest radiograph performed on presentation (n=15), n (%)	
No acute abnormality	8 (53.3)
Unilateral consolidation	2 (13.3)
Bilateral consolidation	2 (13.3)
Blood tests performed on presentation (n=16), n (%)	
Elevated lactate dehydrogenase (n=15)	9 (60)
Elevated C-reactive protein (n=6)	5 (83.3)
Lymphopenia (n=16)	4 (25)
Outcome, n (%)	
Assessed at the emergency department and discharged	6 (33.3)
Admitted to the in-patient ward	12 (66.7)
Prescribed antibiotics	8 (44.4)
Required supplemental oxygen upon admission	4 (22)
Admitted to the intensive care unit	2 (11.1)
Received mechanical ventilation	2 (11.1)
Mortality	0 (0)
Complications, n (%)	
Liver function derangement (n=16)	6 (37.5)
Respiratory failure	3 (16.7)
Acute kidney injury (n=13)	3 (23.1)

[a]High risk was defined as age 65-85 years with any comorbidity or age 49-65 years with chronic lung disease, cardiovascular disease, immunosuppression, diabetes, or hypertension.

Hospitalized Patients

In total, 18 patients with a median age of 54 (range 23-71) years (females: n=10 [55.6%]; high-risk: n=9 [50%]) were assessed in hospital upon virtual ward admission (Table 3). Referral to hospital occurred at a median of 8.5 days (range 3-20 days) since COVID-19 onset. Furthermore, 8 (43.9%) patients were reviewed by a medical officer prior to their virtual ward admission. Among 10 (55%) patients, the most common reason for care escalation was worsening, ongoing, or new-onset dyspnea. In addition, 4 (22%) patients were referred to hospital owing to new or ongoing fever, 3 (17%) for further clinical assessment of chest pain or chest tightness (Table 3), and the remaining 4 (23%) owing to a high symptom burden, functional decline with worsening fatigue, presyncopal symptoms, or suspected bacterial superinfection.

On hospital presentation, 4 (22%) patients had fever, and 4 (22%) had hypoxia and required supplemental oxygen on or shortly after presentation. Of the 18 patients presenting to hospital, 16 (88.9%) most commonly had an elevated lactate dehydrogenase level, liver function derangement, an elevated C-reactive protein level, or lymphopenia on blood tests. Furthermore, 15 (83.3%) patients underwent chest radiography, of which 4 (26.7%) had features of consolidation. Clinically, bacterial pneumonia was suspected in 7 (38.9%) hospitalized patients, and 8 (44.4%) patients were treated with antibiotics. Complications identified during hospitalization included respiratory failure in 3 (16.7%) patients, acute kidney injury in 3 (23.1%) patients, and liver function derangement in 6 (37.5%) patients. Moreover, 2 (11.1%) patients required admission to the ICU for mechanical ventilation.

Several possible predictors of deterioration associated with escalation of care were identified, including the presence of hypertension (OR 3.6, 95% CI 1.28-9.87; *P*=.01), sputum production at symptom onset (OR 5.2, 95% CI 1.74-15.49; *P*=.001), arthralgia at onset (OR 3.8, 95% CI 1.21-11.71; *P*=.02), and an PCR Ct for SARS-CoV-2 RNA of ≤20 on diagnostic NPS (OR 5.0, 95% CI 1.25-19.63; *P*=.02).

Discussion

Principal Findings

This retrospective study describes the characteristics and clinical course of an Australian cohort of patients with COVID-19 treated in a newly established virtual ward. To our knowledge, this is the first study to evaluate a community virtual ward model for patients with COVID-19 and the largest clinical assessment of patients with COVID-19 in Australia to date.

Epidemiologically, most cases were attributed to overseas travel, particularly to the United Kingdom, the United States, or cruise ships, consistent with a previous report from Australia [24]. In this cohort, cough, fever, and headache were the most common symptoms; however, their frequencies differed from those reported in a meta-analysis of hospitalized patients in China [10], thus potentially reflecting a lower disease severity in our cohort or reporting certain differences; our cohort recorded a higher incidence of sore throat (43.5% vs 11.6%) and rhinorrhea (36.8% vs 7.3%) and a lower incidence of chest tightness (9% vs 22.9%) and sputum expectoration (10.8% vs 23.7%) than those reported in the meta-analysis in China [10]. Approximately half of our patients had comorbidities, most commonly including hypertension (n=38, 17%) and asthma (n=24, 10.8%), comparable with a previous report [11].

Our results suggest that a virtual ward model is safe for patients with COVID-19. Overall, 205 (92%) patients recovered without escalation to hospital care. Furthermore, 18 (8.1%) patients required hospital assessment, of which only 12 (5.4%) were admitted to hospital and 2 (0.9%) were admitted to the ICU. This reflects a lower severity of COVID-19 in our cohort (0.9%) compared to that reported previously (5%) [15]. The mortality rate of 0 (0%) in our cohort is consistent with the low nationwide mortality rate of 1.3% in Australia [24], at the time of writing. The median virtual ward admission length was 8 days (range 1-44 days) with a median time to clinical recovery of 16 days (range 10-52 days). This discrepancy may be potentially attributed to returning overseas travelers who acquired the infection abroad. In the context of comorbidity and prior COVID-19–related hospitalization, higher-risk patients stayed longer in the virtual ward. These findings suggest that a virtual model of care can potentially preserve in-patient capacity and resources in hospitals and reduce the risk of disease transmission and hospital-acquired sequelae. Although most of our patients had mild illness, regular monitoring and supportive care may have reduced hospital presentations.

Timely identification of disease exacerbation is imperative for safe virtual care. Several studies have reported high diagnostic agreement between virtual and in-person consultations [25,26]. However, clinicians have cited concerns regarding telephonic consultations, primarily owing to the lack of physical examination [27]. COVID-19 has provided an opportunity to introduce a range of telemedicine approaches to the medical staff, owing to the necessity to deliver safe patient care during a pandemic. Remote assessment of dyspnea, the most common reason for hospital referral in our study, has been challenging during COVID-19 [28]. A difficulty in ruling out urgency upon telephonic consultations may have resulted in a lower threshold for hospital referral in our study, with one-third of hospital-referred patients subsequently discharged after in-person assessment [26]. Our virtual ward model was simple without monitoring equipment beyond household thermometers. Enhanced assessment capabilities through real-time telemonitoring may reduce diagnostic uncertainty [29].

Our hospital-referred patients had a higher median age of 54 (range 14-78) years, being referred to hospital at a median of 8.5 days (range 3-20 days) into their illness. Preexisting hypertension, a proposed risk factor for severe COVID-19, was associated with a 3.6-fold increase in hospitalization rates. Initial symptoms of sputum production and arthralgia were associated with hospital referral, which have not been previously reported. Patients with a PCR Ct for SARS-CoV-2 RNA of ≤20 on diagnostic NPS were 5-fold more likely to be referred to hospital. Although a lower Ct indicates a higher RNA sample quantity, the implication of this value for disease progression is unclear. Further studies are required to validate these findings.

The 2 (0.9%) patients admitted to the ICU had risk factors associated with severe COVID-19 [14]. These 2 men aged >60 years, with preexisting hypertension, presented disease exacerbation on day 10 of their illness and hence required ICU admission and ventilatory support; this is consistent with the reported median of 8-12 days to progression to ARDS and 10 days to ICU admission for patients with severe COVID-19 [14-16,19]. The pathogenesis of this late decline remains unclear; however, pathological hyperinflammation seems likely [30].

Limitations

The limitations of this study include its observational design and retrospective data collection, which resulted in missing data across several variables reported herein. Few patients in our cohort had severe COVID-19. There may have been an ascertainment bias as patients with more severe COVID-19 may have been directly admitted to hospital for the duration of their illness, bypassing the virtual ward.

Conclusions

To our knowledge, this is the largest cohort study of COVID-19 patients in Australia to be described to date and the first to evaluate a virtual ward model of care. This study provides evidence regarding the safety and feasibility of a virtual ward setting to treat patients with COVID-19. Further studies are needed to identify the early predictors of COVID-19 exacerbation and to validate this health care model.

Acknowledgments

We acknowledge all the staff involved in the design, implementation, and delivery of care at the Metro North Virtual Ward. We thank the Metro North Emergency Operations Centre, executive staff, and the Metro North Public Health Unit involved in the COVID-19 response.

Authors' Contributions

OF, EM, and OS collected and analyzed the data and drafted the manuscript. GW and CB contributed to data interpretation and critically revised the manuscript. All authors designed the study, critiqued the manuscript, and approved the final version for submission.

Conflicts of Interest

None declared.

References

1. Rolling updates on coronavirus disease (COVID-19). World Health Organization. 2020 Jul 31. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen [accessed 2020-06-04]
2. Webster P. Virtual health care in the era of COVID-19. Lancet 2020 Apr 11;395(10231):1180-1181 [FREE Full text] [doi: 10.1016/S0140-6736(20)30818-7] [Medline: 32278374]
3. Weinstein RS, Lopez AM, Joseph BA, Erps KA, Holcomb M, Barker GP, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014 Mar;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
4. Flodgren G, Rachas A, Farmer A, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015 Sep 07(9):CD002098 [FREE Full text] [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
5. Khairat S, Meng C, Xu Y, Edson B, Gianforcaro R. Interpreting COVID-19 and Virtual Care Trends: Cohort Study. JMIR Public Health Surveill 2020 Apr 15;6(2):e18811 [FREE Full text] [doi: 10.2196/18811] [Medline: 32252023]
6. Rockwell KL, Gilroy AS. Incorporating telemedicine as part of COVID-19 outbreak response systems. Am J Manag Care 2020 Apr;26(4):147-148 [FREE Full text] [doi: 10.37765/ajmc.2020.42784] [Medline: 32270980]
7. Lam PW, Sehgal P, Andany N, Mubareka S, Simor AE, Ozaldin O, et al. A virtual care program for outpatients diagnosed with COVID-19: a feasibility study. CMAJ Open 2020;8(2):E407-E413 [FREE Full text] [doi: 10.9778/cmajo.20200069] [Medline: 32447283]
8. Lewis C, Moore Z, Doyle F, Martin A, Patton D, Nugent L. A community virtual ward model to support older persons with complex health care and social care needs. Clin Interv Aging 2017;12:985-993 [FREE Full text] [doi: 10.2147/CIA.S130876] [Medline: 28721026]
9. Jones J, Carroll A. Hospital admission avoidance through the introduction of a virtual ward. Br J Community Nurs 2014 Jul;19(7):330-334. [doi: 10.12968/bjcn.2014.19.7.330] [Medline: 25039341]
10. Correction: Clinicopathological characteristics of 8697 patients with COVID-19 in China: a meta-analysis. Fam Med Community Health 2020 May;8(2) [FREE Full text] [doi: 10.1136/fmch-2020-000406corr1] [Medline: 32448784]
11. Hu Y, Sun J, Dai Z, Deng H, Li X, Huang Q, et al. Prevalence and severity of corona virus disease 2019 (COVID-19): A systematic review and meta-analysis. J Clin Virol 2020 Jun;127:104371 [FREE Full text] [doi: 10.1016/j.jcv.2020.104371] [Medline: 32315817]

12. Fu L, Wang B, Yuan T, Chen X, Ao Y, Fitzpatrick T, et al. Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis. J Infect 2020 Jun;80(6):656-665 [FREE Full text] [doi: 10.1016/j.jinf.2020.03.041] [Medline: 32283155]
13. Wu Z, McGoogan J. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
14. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020 May;8(5):475-481 [FREE Full text] [doi: 10.1016/S2213-2600(20)30079-5] [Medline: 32105632]
15. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
16. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
17. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. JAMA Intern Med 2020 Jul 01;180(7):934-943 [FREE Full text] [doi: 10.1001/jamainternmed.2020.0994] [Medline: 32167524]
18. Wang Y, Lu X, Li Y, Chen H, Chen T, Su N, et al. Clinical Course and Outcomes of 344 Intensive Care Patients with COVID-19. Am J Respir Crit Care Med 2020 Jun 01;201(11):1430-1434 [FREE Full text] [doi: 10.1164/rccm.202003-0736LE] [Medline: 32267160]
19. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
20. Richardson S, Hirsch J, Narasimhan M, Crawford JM, McGinn T, Davidson KW, the Northwell COVID-19 Research Consortium, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020 May 26;323(20):2052-2059 [FREE Full text] [doi: 10.1001/jama.2020.6775] [Medline: 32320003]
21. Clerkin KJ, Fried JA, Raikhelkar J, Sayer G, Griffin JM, Masoumi A, et al. COVID-19 and Cardiovascular Disease. Circulation 2020 May 19;141(20):1648-1655. [doi: 10.1161/circulationaha.120.046941]
22. Liang W, Guan W, Chen R, Wang W, Li J, Xu K, et al. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. Lancet Oncol 2020 Mar;21(3):335-337 [FREE Full text] [doi: 10.1016/S1470-2045(20)30096-6] [Medline: 32066541]
23. Minotti C, Tirelli F, Barbieri E, Giaquinto C, Donà D. How is immunosuppressive status affecting children and adults in SARS-CoV-2 infection? A systematic review. J Infect 2020 Jul;81(1):e61-e66 [FREE Full text] [doi: 10.1016/j.jinf.2020.04.026] [Medline: 32335173]
24. Coronavirus Disease 2019 (COVID-19): CDNA National Guidelines for Public Health Units. Australian Government: Department of Health. 2021 Jan 29. URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/cdna-song-novel-coronavirus.htm [accessed 2020-05-13]
25. Martin-Khan M, Wootton R, Whited J, Gray LC. A systematic review of studies concerning observer agreement during medical specialist diagnosis using videoconferencing. J Telemed Telecare 2011;17(7):350-357. [doi: 10.1258/jtt.2011.101113] [Medline: 21983223]
26. Wheeler SQ, Greenberg ME, Mahlmeister L, Wolfe N. Safety of clinical and non-clinical decision makers in telephone triage: a narrative review. J Telemed Telecare 2015 Sep;21(6):305-322. [doi: 10.1177/1357633X15571650] [Medline: 25761468]
27. McKinstry B, Watson P, Pinnock H, Heaney D, Sheikh A. Telephone consulting in primary care: a triangulated qualitative study of patients and providers. Br J Gen Pract 2009 Jun;59(563):e209-e218 [FREE Full text] [doi: 10.3399/bjgp09X420941] [Medline: 19520019]
28. Greenhalgh T, Koh G, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]
29. Angelucci A, Aliverti A. Telemonitoring systems for respiratory patients: technological aspects. Pulmonology 2020;26(4):221-232 [FREE Full text] [doi: 10.1016/j.pulmoe.2019.11.006] [Medline: 31932232]
30. Merad M, Martin JC. Author Correction: Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nat Rev Immunol 2020 Jul;20(7):448 [FREE Full text] [doi: 10.1038/s41577-020-0353-y] [Medline: 32488203]

Abbreviations

ARDS: acute respiratory distress syndrome

CDNA: Communicable Diseases Network Australia
Ct: cycle threshold
ICU: intensive care unit
NPS: nasopharyngeal swab
OR: odds ratio
PCR: polymerase chain reaction

Edited by G Eysenbach; submitted 04.11.20; peer-reviewed by P Yellowlees, X He, A Ugargol; comments to author 05.12.20; revised version received 23.12.20; accepted 10.01.21; published 10.02.21.

Please cite as:
Ferry OR, Moloney EC, Spratt OT, Whiting GFM, Bennett CJ
A Virtual Ward Model of Care for Patients With COVID-19: Retrospective Single-Center Clinical Study
J Med Internet Res 2021;23(2):e25518
URL: http://www.jmir.org/2021/2/e25518/
doi:10.2196/25518
PMID:33529157

Original Paper

Addressing Care Continuity and Quality Challenges in the Management of Hypertension: Case Study of the Private Health Care Sector in Kenya

Aisha Walcott-Bryant[1], PhD; William Ogallo[1], RPh, PhD; Sekou L Remy[1], PhD; Katherine Tryon[2], MBBS, MA; Winnie Shena[3], MSc, CHD; Marloes Bosker-Kibacha[4], BSc, MA

[1]IBM Research Africa, Nairobi, Kenya

[2]CarePay Limited, Nairobi, Kenya

[3]Kenya Obstetrical and Gynecological Society, Nairobi, Kenya

[4]Africa Health Business, Nairobi, Kenya

Corresponding Author:
William Ogallo, RPh, PhD
IBM Research Africa
Catholic University of Eastern Africa
Nairobi,
Kenya
Phone: 254 703023000
Email: william.ogallo@ibm.com

Abstract

Background: Hypertension is a major risk factor of cardiovascular disease and a leading cause of morbidity and mortality globally. In Kenya, the rise of hypertension strains an already stretched health care system that has traditionally focused on the management of infectious diseases. Health care provision in this country remains fragmented, and little is known about the role of health information technology in care coordination. Furthermore, there is a dearth of literature on the experiences, challenges, and solutions for improving the management of hypertension and other noncommunicable diseases in the Kenyan private health care sector.

Objective: The aim of this study is to assess stakeholders' perspectives on the challenges associated with the management of hypertension in the Kenyan private health care sector and to derive recommendations for the design and functionality of a digital health solution for addressing the care continuity and quality challenges in the management of hypertension.

Methods: We conducted a qualitative case study. We collected data using in-depth interviews with 18 care providers and 8 business leads, and direct observations at 18 private health care institutions in Nairobi, Kenya. We analyzed the data thematically to identify the key challenges and recommendations for technology-enabled solutions to support the management of hypertension in the Kenyan private health sector. We subsequently used the generated insights to derive and describe the design and range of functions of a digital health wallet platform for enabling care quality and continuity.

Results: The management of hypertension in the Kenyan private health care sector is characterized by challenges such as high cost of care, limited health care literacy, lack of self-management support, ineffective referral systems, inadequate care provider training, and inadequate regulation. Care providers lack the tools needed to understand their patients' care histories and effectively coordinate efforts to deliver high-quality hypertension care. The proposed digital health platform was designed to support hypertension care coordination and continuity through clinical workflow orchestration, decision support, and patient-mediated data sharing with privacy preservation, auditability, and trust enabled by blockchain technology.

Conclusions: The Kenyan private health care sector faces key challenges that require significant policy, organizational, and infrastructural changes to ensure care quality and continuity in the management of hypertension. Digital health data interoperability solutions are needed to improve hypertension care coordination in the sector. Additional studies should investigate how patients can control the sharing of their data while ensuring that care providers have a holistic view of the patient during any encounter.

(J Med Internet Res 2021;23(2):e18899) doi:10.2196/18899

KEYWORDS

hypertension; health information systems; mobile phone; private sector; Kenya

Introduction

Background

Being the leading cause of morbidity and mortality globally, noncommunicable diseases (NCDs) are a significant global health concern in sub-Saharan Africa and other parts of the world. It is projected that by 2030, deaths from NCDs in sub-Saharan Africa will exceed the combined deaths of communicable and nutritional diseases and maternal and perinatal deaths [1]. The United Nations' sustainable development goal target 3.4 requires that by 2030, all member countries should have reduced premature mortality from NCDs by one-third (baseline 2015) through prevention and treatment and should promote mental health and well-being [2]. However, a 2020 study showed that in most countries, the progress toward achieving this target is slow and that getting back on track requires a combination of prevention, early detection, and treatment within accessible and equitable health systems [3].

This study focuses on hypertension, a major risk factor of cardiovascular disease and one of the most important contributors to NCD-related deaths [4]. By 2015, an estimated 1.13 billion people had hypertension globally [5]. Africa has an estimated hypertension prevalence of 30% among people 15 years or older [6]. This region disproportionately bears the largest burden of the condition, and it is projected that by 2030, over 216 million people in sub-Saharan Africa will have hypertension [7]. In Kenya, the age-standardized prevalence of hypertension is estimated to be 24.5% [8]; however, only 15% are aware of their status and only 8% are on treatment [9]. As there is no cure for hypertension, patients require life-long, continuous care, including screening, access to medication, regular testing, and personalized lifestyle plans.

In response to the impending cardiovascular disease crisis in Kenya, the Kenyan government has developed care guidelines to standardize the management of cardiovascular disease and committed to supporting public and private sector stakeholders to ensure implementation of the guidelines [10]. However, a review of the Kenyan policies on cardiovascular diseases revealed key gaps such as lack of integration with primary health care, lack of private sector involvement, and lack of integrated systems for patient referrals and health data sharing [11]. Furthermore, the Kenyan Ministry of Health identified some critical issues with the current health system in addressing the growing burden of NCDs [12]. These include (1) poor prioritization of NCDs throughout the health care system; (2) poor availability and affordability of safe and efficacious basic technologies and medicines for screening, diagnosis, treatment, and monitoring of NCDs; and (3) inadequate capacity of the health workforce in terms of numbers, equipment, and skills mix for the prevention and control of NCDs.

Objectives

Few studies have investigated the management of hypertension in the Kenyan private health sector in Kenya, despite the sector's fast growth driven by increasing demand for quality health care. It is estimated that up to 42% of the health workers provide services in private facilities [13] and that up to 47% of the poorest Kenyans are seeking care from private facilities [14]. However, there is a dearth of literature on the management of NCDs and associated risk factors in the private health sector of Kenya. Little is known about the experiences of care providers in this sector and the challenges they face and how health information systems can be used to address these challenges.

We carried out a case study that aims to understand the management of hypertension in the private health sector in Nairobi, Kenya. The objectives of this study are two-fold. The first objective is to identify and characterize the challenges in the management of hypertension as perceived by health care professionals in the private health sector. The second objective is to derive recommendations for the design and functionality of a digital health information exchange and decision support platform for addressing care continuity and quality challenges in the management of hypertension and other chronic illnesses. This paper reports our findings.

Methods

Study Design

This descriptive case study used a multimethod approach involving in-depth stakeholder interviews and direct observations. This study was conducted at 18 health care institutions located within the greater Nairobi area of Kenya and providing care to patients of low-, middle-, and high-socioeconomic status. These included 6 pharmacies, 6 general health clinics, and 6 specialist outpatient clinics associated with hospitals. The chosen institutions were purposefully selected by the study team such that each socioeconomic stratum was represented by 2 pharmacies, 2 general clinics, and 2 specialist clinics. The study was designed by a team consisting of domain experts from different disciplines, including health care, user interaction and design, and computer science. To mitigate bias, questions were mainly focused on the industry and not specific institutions, and all responses were anonymized.

Data Collection

Data were collected by 5 nurses who were recruited through the National Nurses Association of Kenya and trained by the study team. Of the 5 nurses, 3 collected data from 4 facilities, and 2 collected data from 3 facilities, giving a total of 18 facilities. All the institutional visits were carried out for the purpose of research data collection only, and none of the 5 data collectors had affiliations with any of the facilities studied.

In each location, 4 key processes were carried out for this qualitative study:

1. In-depth interviews: an in-depth interview was conducted with the lead health care professional at the location to understand their education, training, and experience. In

addition, the interviews aimed to understand the health care professional's management of patients with hypertension, tools available to them in their workplace, and the challenges they face in managing patients with hypertension.

2. Roleplay: a roleplay exercise was conducted with the lead health care professional to observe their routine management and care of (1) a patient newly diagnosed with hypertension and (2) patients with hypertension requiring a routine.
3. Facility tour: a facility tour was conducted to identify and understand what resources were available within the facility to serve patients with hypertension.
4. Contextual observations: observations were recorded about the location to understand the context in which the health professionals work and where patients visit.

In addition, interviews were conducted with the business leads of 8 institutions to understand their perspectives on the increasing prevalence of NCDs and to understand the management through their institution. For each location, detailed notes were taken against the interview questions, role-play guides, tour guides, and observation guides.

Data Analysis

All research notes were collated and manually analyzed using the thematic analysis guide by Braun and Clarke [15]. The goal of the analysis was to discover themes about the challenges and gaps associated with the management of hypertension at the study sites. Specifically, the analysis aimed at exploring participants' perspectives on how they managed suspected and confirmed cases of hypertension and why this process was challenging. Clusters of semantically similar features and responses were first identified from the transcribed research notes. Next, the specific potential themes and subthemes were identified from the clusters. Subsequently, the potential themes and subthemes were iteratively verified against the original data. Finally, the final themes and subthemes were defined, refined, and reported. These themes and subthemes were used to inform the design and development of a novel platform to support the management of NCDs, provide continuity of care across facilities and systems, and improve the quality of care.

Results

Hypertension Management Challenges in the Kenyan Private Health Care Sector

Through this study, we identified 6 themes and 17 subthemes regarding the challenges in the management of hypertension in the private health care sector in Kenya. These are summarized in Table 1 and discussed below.

Table 1. Themes and subthemes about the management of hypertension in the private health care sector in Kenya.

Theme	Subtheme
High cost of hypertension management	1. Limited resources introducing priorities that compete with hypertension management 2. High cost of medications prevents effective management of hypertension 3. Prescribing habits and patient preferences fail to mitigate the cost of medications
Limited health literacy	1. Ignorance about hypertension prevents effective management of the disease 2. Need to address health literacy and improve hypertension awareness
Lack of self-management support	1. Patients lack access to their hypertension management data 2. Patients lack access to postappointment information for self-management 3. Lack of tailoring of advice to patients due to limited capacity to monitor patients in the community 4. Limited off-site follow-up and outreach of patients in the community
Ineffective referral systems	1. Patients perceive self-referral to higher-level hospitals as acceptable 2. Knowledge and skill gaps exist at the different levels of the health care pyramid 3. There is no financial incentive for private hospitals to refer patients back to the community or lower levels of care 4. Litigation fears around task shifting hinder downward referrals to community hospitals
Inadequate care provider training	1. Lack of adequate training on the management of hypertension 2. Variability in decision-making aids
Inadequate regulation	1. Lack of adequate regulations on the quality of services in private health facilities 2. High number of unregulated pharmacies

High Cost of Hypertension Management

We found that the cost of a hypertension consultation at a private clinic or hospital in the greater Nairobi area ranges from US $4 to US $40. Patients are encouraged to attend a monthly consultation for progress monitoring and renewal of prescriptions. Most payments are out of pocket; thus, the cost of care affects the frequency of consultations. Some respondents stated that consultation fees place a significant burden on patients with lower income levels. Often, health care budget burdens compete with other priorities such as housing and education (subtheme 1):

> *Health is just not a priority for our clients, they [patients] are worrying about education and housing,*

> *not health. They come in when they absolutely have to.* [Healthcare Group General Manager]
>
> *It is expensive for people to keep going back to the hospital, so they stop going and either stop their medication or keep buying it without a checkup.* [Pharmacy CEO]

An additional financial issue in the management of hypertension is the cost of medication (subtheme 2). Seven of the 18 health care professionals interviewed stated that medication cost was one of the biggest factors preventing their patients from effectively managing their hypertension. In addition, some respondents noted that habits such as prescribing branded drugs fail to mitigate the costs of medications (subtheme 3):

> *Doctors write branded prescriptions and patients insist on receiving medication that is written on the prescription, even if there is a cheaper generic version available.* [Pharmacy CEO]

Limited Health Literacy

Nine of the 18 health care professionals interviewed identified limited knowledge about hypertension and its management as the primary factor preventing their patients from effective management of their condition (subtheme 4):

> *Ignorance in the market is the biggest issue in managing chronic disease.* [Clinic General Manager]
>
> *No one understands blood pressure, what it means and how to manage it.* [Pharmacy CEO]
>
> *If you try and screen patients, they think you are over investigating them for more money.* [Clinic General Manager]

Respondents recognized the need to address the health care literacy of their patients to improve hypertension management (subtheme 5). For example, most interviewed pharmacists identified their role as critical for patient education and follow-up. They also identified specific challenges within their role in patient education. These included (1) limited access to patient health history, (2) lack of education and training in the management of NCDs, and (3) potential conflicts of interest in selling medications. Other approaches for improving health literacy included learning from experiences in the management of other chronic illnesses such as HIV/AIDS and educating patients during routine care visits:

> *We need to learn from the HIV work on how to get the prevention message out.* [Hospital CEO]
>
> *We offer health talks in companies for our capitated patients as it is worth our while doing that.* [Healthcare Group General Manager]

Lack of Self-Management Support

This study identified several gaps pertaining to tools for supporting patients with hypertension. First, patients lack easy access to their hypertension management health records (subtheme 6). During the interviews, the health care professionals acknowledged that some patients used apps and notebooks to record blood pressure measurements. Most patients can obtain printed copies of their care notes on request from the institution. However, one institution forbade access to health records by patients:

> *Records are given [to patients] by word of mouth and files not issued out due to privacy purposes.* [Hospital CEO]

Second, the health care professionals interviewed identified the incompatibility of health records between institutions as another issue. Specifically, the respondents noted that the transfer of records between institutions is a complicated process. As a result, important details of patient cases can be missed and investigations can be repeated, leading to additional medical costs for patients and their insurers.

Third, patients lack information for supporting self-management once they leave the doctor's office (subtheme 7). Health care professionals counseled their patients on lifestyle and medication adherence at meetings or appointments. However, there were very few take-home materials for their patients. Of the 18 institutions interviewed, only 2 had leaflets or pamphlets on hypertension. The fourth gap is on counseling, where generic advice on lifestyle changes and medication adherence was given rather than personalized advice (subtheme 8). Interviewees noted that they see a limited view of their patients' lives, and there is little monitoring of patients in the community.

The final gap identified was limited to follow-up and outreach with patients in the community (subtheme 9). One hospital had phone call outreach and counseling, one clinic had a monthly health education session for patients with hypertension, and one health care group had health talks for their capitated groups. In total, 15 out of 18 institutions relied entirely on appointments to educate and assist patients with the management of their condition:

> *There is no active outreach into the community; the hospital waits for patients to come back for an appointment.* [Hospital CEO]
>
> *Many patients fall through the cracks and only come back with serious consequences of their condition.* [Hospital CEO]

Ineffective Referral Systems

The Kenyan health care system is organized in a *pyramid of care*, where less complex cases should be managed in a community or primary care setting and increasingly complex cases should be referred to general or specialist hospitals. We found that more complex cases, with greater levels of comorbidities and complications, were seen in hospitals rather than clinics. However, the interviewees acknowledged that many uncomplicated or well-controlled cases are also managed in hospitals, thereby resulting in higher costs and inefficient use of resources. The reasons why many patients receive care in hospitals, rather than in clinics, are discussed below. First, many patients perceive self-referrals to higher-level hospitals as acceptable (subtheme 10). We found that patients bypass lower-level facilities because they think that they would be referred to a higher-level facility and that it is easier and more cost-effective to go straight to higher-level institutions:

> *Patients realize that lower-cost options often result in referrals costing more in health, time and money, so they prefer to come straight to hospital.* [Hospital CEO]
>
> *It is not acceptable in the market to see a clinical officer or a nurse manage your condition, people expect to see a medical officer.* [Healthcare Group CEO]
>
> *Patients think that more expensive is better health care and we do not change that perception, so they keep coming back to us for routine care.* [Hospital CEO]
>
> *It is about giving patients what they want.* [Healthcare Group CEO]

Second, there are skill gaps in the health care workforce at different levels of the care pyramid. For example, respondents stated that there are care providers at lower-level hospitals who do not have the skills necessary for managing hypertension. Other respondents identified the lack of a specialized skilled health workforce at higher-level hospitals as a significant barrier to the management of hypertension:

> *There is no truly skilled primary care in the community to manage NCDs.* [Hospital CEO]
>
> *Community health workers screen in the community, but patients have to come to the clinic for follow up.* [Clinic General Manager]
>
> *We have a significant shortage of skilled staff at the subspecialty level in chronic disease.* [Healthcare Group General Manager]

Third, it was broadly acknowledged by interviewees that there is no financial incentive for private hospitals to refer patients with hypertension back to the community- or lower-level facilities, even if they are well managed (subtheme 12). Hospitals themselves have little incentive to task shift within their institutions and to use lower cadres of professionals in the health care system as insurers to date are not generally willing to accept capitated services and few providers offer capitated services to their customers. Capitation refers to a payment contract between an insurer and a care service provider in which the insurer periodically pays the service provider a fixed amount for an enrolled person regardless of whether or not the patient seeks care. Respondents opined that the lack of capitation discourages task shifting, possibly because of the associated hospital revenue losses:

> *Insurers are not yet putting pressure on us to reduce the cost of care.* [Hospital CEO]
>
> *Insurers have not historically accepted capitated rates for their customers in return for a lock-in to [named health care group] services, so we have no incentive to task shift.* [Healthcare Group CEO]

Fourth, the fear of litigation around task shifting hinders the referral of patients seen at higher-level facilities back to the community (subtheme 13). Such fears are associated with the lack of policies governing referrals from higher-level hospitals back to lower-level facilities:

> *There is some pressure from NHIF and insurers to reduce the cost of care, but we still have doctor-led service due to litigation worries about task shifting to lower cadres of professionals.* [Hospital CEO]
>
> *We would wait for policy change at the government level before risking task shifting.* [Hospital CEO]

Inadequate Care Provider Training

This study highlighted specific concerns regarding inadequate training on NCDs and lack of continued professional development (CPD; subtheme 14). For example, 8 of the 18 health care professionals interviewed took no additional training in hypertension management after their initial qualification. Of the remaining 19 who had some training, 4 received training via CPD organized by pharmaceutical companies and focused on companies' branded medications. This situation was even more severe for pharmacy staff interviewed, with 4 of 6 staff having no further training following qualification:

> *Most CPD is provided by pharmaceutical companies trying to promote their medication, and there is a critical need for independent CPD provision.* [Hospital CEO]
>
> *CPD should be regulated by separate boards from those providing the CPD. Currently, many of the professional bodies both provide and regulate CPD.* [Hospital CEO]
>
> *CPD is not obligatory for healthcare professionals. In the case of re-licensing, most [professionals] qualify by completing their required hours. So many do not continue to learn after they have completed their initial training. This is a particular issue for NCDs as little of their initial training would have covered these at a time when HIV, TB, and Malaria were the main focus.* [Hospital CEO]

Our research also identified a large variability in the decision-making aids used by different health care professionals to guide the care of their patients (subtheme 15). For example, of the 18 institutions interviewed, each used different sets of guidelines to manage their patients with hypertension. This highlighted not only the inconsistency in training across the board but also the challenges faced by patients moving between institutions that follow different guidelines.

Inadequate Regulation

During the interviews, there were important regulatory concerns that were identified. First, although health care professionals in private facilities are regulated by their respective professional bodies, the facilities themselves are not regulated for quality of service (subtheme 16):

> *There is no quality inspection of private facilities, so as long as you pass financial audit you can stay open regardless of the quality of care. An independent quality commission is much needed.* [Hospital CEO]
>
> *We worry about bad newspaper articles that will stop people coming to our hospital, not a quality audit. We need to keep our patients happy.* [Hospital CEO]

Second, the Pharmaceuticals and Poisons Board has been tasked with the regulation of the pharmacy industry in Kenya. However, there remain significant issues and a large number of unregulated pharmacies across the country (subtheme 17):

> *Some wholesalers insist on seeing licensing before they sell pharmaceuticals, but others do not, so you can still sell pharmaceuticals without a license.* [Pharmacy CEO]
>
> *The Pharmacy and Poisons Board announce their investigations the week before they do them so many pharmacies are closed when they come to investigate.* [Pharmacy CEO]
>
> *Patients do not know who is legitimate and who is not. No one explained what the green cross system meant.* [Pharmacy CEO]

Proposed Approach to Addressing Care Continuity and Quality Challenges

A total of 22 recommendations were identified from the solutions described by the interviewees in this study (Table 2). Of these solutions, 6 directly involved the use of health information systems to address different aspects of care: (1) reducing costs of hypertension management (recommendation 3), (2) supporting continuous patient follow-up and education (recommendation 10), (3) enabling self-measurements (recommendation 15), (4) enhancing patient adherence (recommendation 16), (5) supporting interoperability for health facilities and systems to support the health care pyramid (recommendation 17), and (6) adoption of local clinical guidelines (recommendation 18).

Table 2. Solutions for improving management of hypertension recommended by interviewees.

Theme	Recommendations
High cost of hypertension management	1. More affordable or subsidized medications 2. Greater penetration of insurance coverage 3. Technology solutions to reduce the cost of care, for example, telemedicine, self-service machines, monitoring and triaging patients, and advising patients (eg, on generic medications) 4. Financial incentives for technology use over in-person visits for health care professionals or organizations 5. Incentive schemes for individuals to make lifestyle changes and be compliant with medications 6. Improved infrastructure so that people can walk more frequently and more safely 7. National policies around healthy food 8. Improved financial access to healthy food
Limited health literacy	1. Use of nurse educators 2. Technology to support continuous patient follow-up and counseling 3. Education materials, including videos in the waiting bays in clinics or hospitals, and materials for the general public 4. Increasing the role of pharmacists in patient education 5. Educating children in school about health and healthy lifestyle
Lack of self-management support	1. Introduction of patient held records 2. Kiosks for patients to measure themselves independently 3. Tracking devices for patient lifestyle and medication adherence monitoring in the community
Ineffective referral systems	1. Vertical integration between different tiers of care delivery through efforts such as enabling communication and cooperation between care providers, improving referral care coordination and supervision, and providing transportation
Inadequate care provider training	1. Adoption of local guidelines 2. Training of pharmacy staff on the management of NCDs[a] 3. Independent CPD[b] for health care professionals 4. Compulsory CPD linked to licensing
Inadequate regulation	1. Greater regulation of health care organizations (including pharmacies) on quality

[a]NCD: noncommunicable disease.

[b]CPD: continued professional development.

The themes and recommendations identified by our research suggest that the inherent complexities of Kenya's health care landscape make it difficult for stakeholders to ascertain a deep understanding of the patient's condition and effectively coordinate efforts to deliver high-quality care. This reveals the overarching lack of capacity for enabling care quality and continuity within a fragmented ecosystem. In this regard, we identified the need to strengthen health systems in the Kenyan private and public sectors to support the management of hypertension and to ensure care continuity and quality within and across facilities. On the basis of the lessons learned from our research, we proposed a patient-centered platform, called the *Digital Health Wallet Platform*, which aims to improve the documentation and exchange of health data gathered during care provision by allowing patient-consented and mediated data

sharing [16]. The platform is built on top of a blockchain network to ensure privacy and auditability of care provision.

This platform has 3 core components: a care management platform, a patient care wallet, and a clinical encounters app (Table 3). The *care management platform* manages clinical encounters and enables the patient-controlled exchange of health data via blockchain protocols while adhering to the Fast Healthcare Interoperability Resources (FHIR) version 3.0 standard [16]. It supports data sharing and guideline-based decision making in outpatient clinical workflows. The *patient care wallet* is a mobile app where patients view their health records and can track who accessed their records [16]. It also enables patient-consented data sharing via a decentralized consent management blockchain system. Finally, the *clinical encounters app* supports care providers through workflows and enables requests for patient consent to share health data between providers [16].

The architecture of the proposed platform is illustrated in Figure 1 and has been described in detail by Osebe et al [16]. End users access the system via specific apps on smartphones (for patients) and tablets (for providers). Patients use a *patient care wallet* to view their health records, view who accessed records, consent to the sharing of their health data, and communicate with their care providers via text messages. Care providers use a *clinical encounters app* to document care observations and interventions made during encounters in a clinical workflow, view patient records, and initiate patient-consented data sharing processes.

Table 3. Key components and capabilities of the proposed digital health wallet platform.

Component	Key capabilities	Targeted recommendation
Care management platform	1. Mediation of patient-consented health data sharing 2. Blockchain-managed execution of clinical workflows 3. Interoperability across disconnected systems 4. Adheres to FHIR[a] version 3.0 standard	• Supporting interoperability (recommendation 17) • Reducing costs of care (recommendation 3)
Patient care wallet	1. Enables patients to view their health records 2. Enables patients to view who accessed records 3. Enables patients to consent to share health data 4. Support patient-provider outreach conversations	• Follow-up and education (recommendation 10) • Enhancing patient adherence (recommendation 16)
Clinical encounters app	1. Enables documentation of care by providers in clinical workflows 2. Enables requests for consent from patients 3. Ensures holistic view of the patient	• Adoption of local guidelines (recommendation 18) • Enabling self-measurements (recommendation 15)

[a]FHIR: Fast Healthcare Interoperability Resources.

The *application services layer* consists of several components that expose core functionalities to end users. The *application service* submits care observations, personal health data, and outreach conversations generated by end users to an off-ledger database. It also generates metadata that capture events and user actions on medical records. The *care guideline service, patient outreach service, anomalous patterns of care service,* and *transition of care summarization service* perform different functions to provide data viewership, insights, patient summaries, and care recommendations intended to support decision making at the point of care.

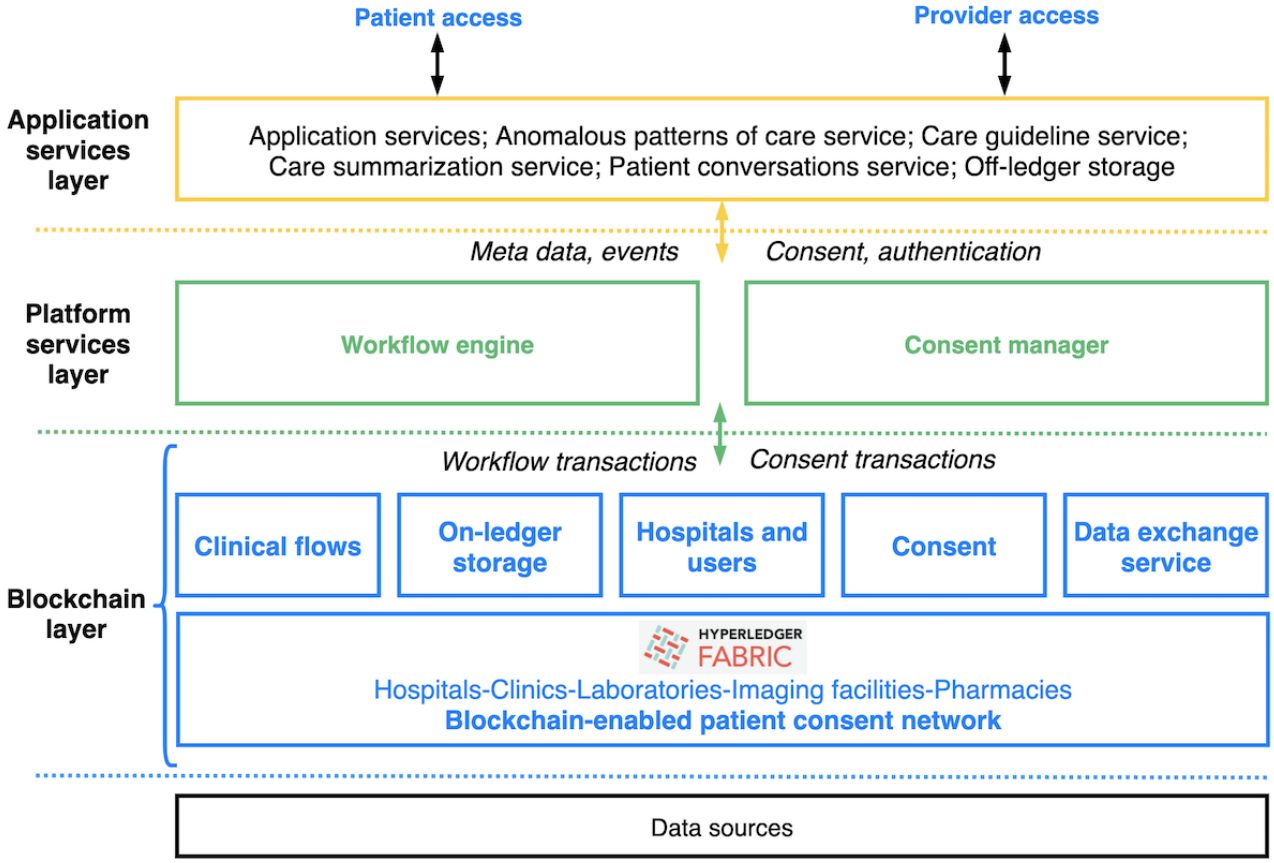
The *platform services layer* links the application layer and the blockchain layer to ensure access to blockchain services by other services. This layer consists of a *workflow engine* that orchestrates clinical workflows. Workflows consist of interaction points (encounters) such as triage, consultation, and laboratory investigation. These interaction points can be encoded as chain codes that can be controlled and executed in predefined sequences while specifying the actors, actions, inputs, and outputs pertaining to the workflow. The *Consent Manager* enforces privacy preservation and data protection by ensuring that access and viewership of data in the system is controlled by patients. It empowers patients to give explicit consent to the sharing of their clinical encounters data across organizations. Only authorized users who are onboarded into the system can initiate data sharing processes, and no encounter data can be accessed across separate organizations without valid patient consent.

The *blockchain layer* ensures decentralized trust among participating organizations and patients. It uses smart contracts to map out all encounters in a clinical workflow and ensures that only users with permission can access the encounters and associated data. The layer has an *on-ledger storage* that stores only metadata, encryption keys, and links to data in an off-ledger data store. This ensures the preservation of sensitive patient data that remain stored in off-ledger electronic medical records on-site or ideally in a Health Insurance Portability and Accountability Act (HIPAA)–compliant cloud. The blockchain layer also has an Organizations and Users component for onboarding actors into the system and enforcing access control levels and rules that define users who can create, read, update, or share data in the system.

Such a platform provides value to health care stakeholders in managing not only hypertension but also other diseases, particularly as it takes a patient-centric approach. For patients, the platform would enable easy patient-consented sharing of health data while preserving privacy, better patient engagement, and the potential to monetize personal health records. The availability of clinical and personal health data would all be enabled on mobile devices. For care providers, the system would provide holistic views of patient histories, leading to better

diagnoses and treatments. Furthermore, efficient access to patient information would optimize care coordination among care providers. Payers would experience reductions in administrative delays and costs associated with paper records and fragmented systems and a reduction in costs associated with unnecessary or duplicate tests and prescriptions and ease auditability of care provision across health facilities.

Figure 1. High-level overview of the proposed digital health wallet platform for enabling care continuity.



Discussion

Principal Findings

In this study, we investigated stakeholders' perspectives on the management of hypertension in the private health sector in Kenya and described the design and functionality of a novel digital health solution for addressing the care continuity and quality challenges in the management of hypertension in the sector. Our findings suggest that the Kenyan private health care sector is a complex, fragmented ecosystem with several challenges such as high cost of care, limited health care literacy, lack of self-management support, and ineffective referral systems. We established that the sector needs health data sharing and decision-making platforms to support care coordination and continuity while acknowledging the fact that addressing all the challenges requires a multifaceted approach.

Interestingly, one of the most significant challenges identified in our study was the lack of sufficient vertical integration between hierarchical tiers of health care delivery. This makes it difficult for practitioners to have a complete understanding of the patient and effectively coordinate efforts to deliver high-quality hypertension care. Appropriate triage and referral from lower-level to higher-level health care institutions is critical to the efficient use of limited resources. Less complex cases should be managed in a community or primary care setting, and increasingly complex cases should be referred to higher-level hospitals. Kenya has articulated this strategy in its health sector referral implementation guidelines [17]. However, the lack of communication and visibility of data collected in clinical encounters across different care facilities makes the referral process unreliable and inefficient. This is true not only for Kenya but also for other countries. For example, in the United States, up to 50% of referrals have no communication between care providers and up to 70% have incomplete documentation [18]. Although there is a dearth of literature on referral practices in low- and middle-income countries, such settings are more likely to experience more difficult referral communication challenges due to higher fragmentation and more scarce resources. Consequently, digital health solutions adopted in these settings should enable patient-consented data sharing while ensuring that care providers have the necessary tools for having a holistic view of longitudinal patient records siloed in separate facilities.

Comparison With Related Work

Our findings on the health literacy of patients and on the knowledge and skill gaps among care providers mirror the findings of the Healthy Heart study [19]. In the Healthy Heart study [19], only 30% of the interviewed patients knew that heart attacks are associated with high blood pressure, and of those

who had heard of hypertension, only 18% identified salt as a risk factor and only 33% identified medication as a method for reducing blood pressure. Furthermore, less than 40% of health care professionals at clinics and dispensaries identified lifestyle changes as methods to address hypertension and less than 75% identified medication as a method to treat hypertension. To improve hypertension health literacy, knowledge, and skills among patients and care providers, policy and organizational approaches such as training and continuous medical education are required. In addition, digital health solutions, such as the one proposed in this study, should have capabilities that support guideline-based decision-making and patient engagement and self-management.

Overall, Kenya can learn from its significant experience in tackling the HIV epidemic. The key themes identified in this study have been addressed, to some extent, in the ongoing management of HIV in Kenya. For example, the study by Oduor et al [20] suggests that the use of technology to manage HIV is influenced by the roles and routines of the patients and clinicians, and such use can inform the design of technologies that can support patients living with comorbid HIV and hypertension. However, directly importing the HIV model for additional conditions would be a challenge, both due to the overall cost of the model (HIV accounts for over 35% of total health expenditure in Kenya) and funding sources (over 70% of HIV care is funded by international organizations). Solutions must be identified for hypertension and other NCDs that learn from the experiences of HIV management and seek more cost-effective mechanisms of managing these conditions. The solutions should be patient-centered and seek to improve health outcomes, as exemplified in the widely adopted chronic care model [21]. Solution frameworks such as the Digital Health Wallet Platform proposed in this study present valuable approaches for enabling care continuity and quality within fragmented health care systems. Such solutions can be used to improve the efficiency of care coordination and patient referrals by supporting communication between care providers, eliminating unnecessary paperwork, and reducing duplicate services. Such solutions are useful for promoting patient engagement; enhancing medication adherence, safety, and effectiveness; and empowering patients with the management of their health data.

The usability and feasibility of digital health solutions for hypertension and other NCDs should be adequately investigated while considering the important barriers such as low digital literacy among users, high costs of implementation and maintenance, unreliable infrastructure, and weak regulatory frameworks. On the basis of the findings of this study, we have developed the Digital Health Wallet Platform [16] and are in the process of testing its feasibility for the clinical management of HIV and comorbid hypertension and/or diabetes across 3 clinical workflows. We have also identified new research directions that are complementary to the platform, which are important for improving health care delivery in settings such as Kenya. These areas, which are critical for supporting clinical encounter analytics, include context-based abstractive text summarization [22] and the detection of anomalous patterns of care [23,24] and are actively being pursued by our team.

Conclusions

This study identified key challenges faced by care providers managing hypertension in the Kenyan private health care sector. Our findings suggest that significant policy, organizational, and infrastructural changes are required to address these challenges and ensure care quality and continuity. Multiple partners, including health care payers, health care providers, health care regulators, and technology partners, are required to come together in this endeavor. Additional studies should investigate how digital health data interoperability can be achieved within fragmented ecosystems such as the private sector in health care in Kenya by enabling patients to control and manage the sharing of their data while ensuring that care providers have a holistic view of the patient during any encounter.

Conflicts of Interest

None declared.

References

1. Global action plan for the prevention and control of NCDs 2013-2020. In: WHO. Geneva, Switzerland: World Health Organization; 2013.
2. Transforming our world: the 2030 agenda for sustainable development. United Nations General Assembly. 2015. URL: https://sustainabledevelopment.un.org/post2015/transformingourworld [accessed 2021-01-06]
3. NCD Countdown 2030 collaborators. NCD Countdown 2030: pathways to achieving Sustainable Development Goal target 3.4. Lancet 2020 Sep 26;396(10255):918-934 [FREE Full text] [doi: 10.1016/S0140-6736(20)31761-X] [Medline: 32891217]
4. Lawes CMM, Vander Hoorn S, Law MR, Elliott P, MacMahon S, Rodgers A. Blood pressure and the global burden of disease 2000. Part II: estimates of attributable burden. J Hypertens 2006 Mar;24(3):423-430. [doi: 10.1097/01.hjh.0000209973.67746.f0] [Medline: 16467640]
5. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19·1 million participants. Lancet 2017 Dec 07;389(10064):37-55 [FREE Full text] [doi: 10.1016/S0140-6736(16)31919-5] [Medline: 27863813]
6. Ataklte F, Erqou S, Kaptoge S, Taye B, Echouffo-Tcheugui JB, Kengne AP. Burden of undiagnosed hypertension in sub-saharan Africa: a systematic review and meta-analysis. Hypertension 2015 Feb;65(2):291-298. [doi: 10.1161/HYPERTENSIONAHA.114.04394] [Medline: 25385758]

7. Adeloye D, Basquill C. Estimating the prevalence and awareness rates of hypertension in Africa: a systematic analysis. PLoS One 2014;9(8):e104300 [FREE Full text] [doi: 10.1371/journal.pone.0104300] [Medline: 25090232]
8. Mohamed SF, Mutua MK, Wamai R, Wekesah F, Haregu T, Juma P, et al. Prevalence, awareness, treatment and control of hypertension and their determinants: results from a national survey in Kenya. BMC Public Health 2018 Nov 07;18(Suppl 3):1219 [FREE Full text] [doi: 10.1186/s12889-018-6052-y] [Medline: 30400858]
9. Kenya STEPwise Survey for non-communicable diseases risk factors 2015 report. Kenya Ministry of Health.: Division of Non-Communicable Diseases; 2018. URL: https://www.health.go.ke/wp-content/uploads/2016/04/Steps-Report-NCD-2015.pdf [accessed 2021-01-06]
10. Kenya national guidelines for cardiovascular diseases management. Kenya Ministry of Health.: Division of Non-communicable Diseases; 2018. URL: ttps://www.health.go.ke/wp-content/uploads/2018/06/Cardiovascular-guidelines-2018_A4_Final.pdf [accessed 2021-01-06]
11. Asiki G, Shao S, Wainana C, Khayeka-Wandabwa C, Haregu TN, Juma PA, et al. Policy environment for prevention, control and management of cardiovascular diseases in primary health care in Kenya. BMC Health Serv Res 2018 May 09;18(1):344 [FREE Full text] [doi: 10.1186/s12913-018-3152-4] [Medline: 29743083]
12. Kenya national strategy for the prevention and control of NCDs 2015 - 2020. Kenya Ministry of Health.: Division of Non-communicable Diseases; 2015. URL: https://ilakenya.org/kenya-national-strategy-for-the-prevention-and-control-of-ncds-2015-2020/ [accessed 2021-01-06]
13. Kenya health service delivery indicator survey 2018 report. The World Bank and the Government of Kenya. 2019. URL: https://ncpd.go.ke/wp-content/uploads/2020/05/Final-KESDI-Health-Technical-Report-1.pdf [accessed 2021-01-06]
14. Healthcare in Kenya and opportunities in the Kenya health sector. Alliance Experts. 2017. URL: https://www.allianceexperts.com/en/knowledge/countries/africa/the-healthcare-sector-in-kenya/ [accessed 2021-01-06]
15. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res in Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
16. Osebe S, Walcott-Bryant A, Weldemariam K, Wachira C, Matu F, Bore N, et al. Enabling Care Continuity using a Digital Health Wallet. In: 2019 IEEE International Conference.: IEEE; 2019 Jun Presented at: 2019 IEEE International Conference on Healthcare Informatics (ICHI); 10-13 June 2019; Xi'an, China URL: https://ieeexplore.ieee.org/document/8904625
17. Kenya health sector referral implementation guidelines 2014. Kenya Ministry of Health.: Division of Emergency and Disaster Risk; 2014. URL: http://publications.universalhealth2030.org/uploads/ministry-of-health-referral-guidelines.pdf [accessed 2021-01-06]
18. Mehrotra A, Forrest CB, Lin CY. Dropping the baton: specialty referrals in the United States. Milbank Q 2011 Mar;89(1):39-68 [FREE Full text] [doi: 10.1111/j.1468-0009.2011.00619.x] [Medline: 21418312]
19. Ogola EN. Healthy Heart Africa: the Kenyan experience. In: PASCAR Hypertension Task Force Meeting. 2015 Aug Presented at: PASCAR Hypertension Task Force Meeting; August, 2015; London, UK URL: https://www.pascar.org/uploads/files/HEALTHY_HEART_AFRICA_Elijah_Ogola.pdf
20. Oduor E, Pang C, Wachira C, Bellamy RKE, Nyota T, Remy SL, et al. Exploring Rural Community Practices in HIV Management for the Design of Technology for Hypertensive Patients Living with HIV. In: Proceedings of the 2019 on Designing Interactive Systems Conference. New York, NY, United States: Association for Computing Machinery; 2019 Jun Presented at: DIS '19: Designing Interactive Systems Conference 2019; June, 2019; San Diego CA USA p. 1595-1606 URL: https://dl.acm.org/doi/10.1145/3322276.3322348
21. Wagner E. Chronic disease management: what will it take to improve care for chronic illness? Eff Clin Pract 1998;1(1):2-4. [Medline: 10345255]
22. Cintas C, Ogallo W, Walcott-Bryant A, Remy SL, Akinwande V, Osebe S. Towards neural abstractive clinical trial text summarization with sequence to sequence models. In: 2019 IEEE International Conference on Healthcare Informatics (ICHI).: IEEE; 2019 Jun Presented at: 2019 IEEE International Conference on Healthcare Informatics (ICHI); 10-13 June 2019; Xi'an, China URL: https://ieeexplore.ieee.org/document/8904526/
23. Neill DB. Using artificial intelligence to improve hospital inpatient care. IEEE Intell. Syst 2013 Mar;28(2):92-95. [doi: 10.1109/mis.2013.51]
24. Speakman S, Somanchi S, McFowland III E, Neill DB. Penalized Fast Subset Scanning. J of Computational and Graph Stati 2016 May 10;25(2):382-404. [doi: 10.1080/10618600.2015.1029578]

Abbreviations

CPD: continued professional development
FHIR: Fast Healthcare Interoperability Resources
NCD: noncommunicable disease

Edited by G Eysenbach; submitted 26.03.20; peer-reviewed by C Mavrot, P Banik; comments to author 18.07.20; revised version received 30.09.20; accepted 19.10.20; published 17.02.21.

Please cite as:
Walcott-Bryant A, Ogallo W, Remy SL, Tryon K, Shena W, Bosker-Kibacha M
Addressing Care Continuity and Quality Challenges in the Management of Hypertension: Case Study of the Private Health Care Sector in Kenya
J Med Internet Res 2021;23(2):e18899
URL: http://www.jmir.org/2021/2/e18899/
doi:10.2196/18899
PMID:33595446

Original Paper

Adoption of Electronic Health Records (EHRs) in China During the Past 10 Years: Consecutive Survey Data Analysis and Comparison of Sino-American Challenges and Experiences

Jun Liang[1], MS; Ying Li[2], MD; Zhongan Zhang[3], BS; Dongxia Shen[4], MS; Jie Xu[1], MS; Xu Zheng[5], MSc; Tong Wang[6], MS; Buzhou Tang[7], PhD; Jianbo Lei[5,8,9], MD, PhD; Jiajie Zhang[10], PhD

[1]IT Center, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

[2]Department of Burns and Plastic Surgery, The Affiliated Hospital of Southwest Medical University, Luzhou, China

[3]Performance Management Department, Qingdao Central Hospital, Qingdao, China

[4]Editorial Department, Journal of Practical Oncology, Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

[5]Center for Medical Informatics, Peking University Third Hospital, Beijing, China

[6]School of Public Health, Jilin University, Changchun, China

[7]Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China

[8]Institute of Medical Technology, Health Science Center, Peking University, Beijing, China

[9]School of Medical Informatics and Engineering, Southwest Medical University, Luzhou, China

[10]School of Biomedical Informatics, University of Texas Health Sciences Center, Houston, TX, United States

Corresponding Author:
Jianbo Lei, MD, PhD
Institute of Medical Technology
Health Science Center
Peking University
38 Xueyuan Rd
Haidian District
Beijing, 100191
China
Phone: 86 (10) 8280 59
Email: jblei@hsc.pku.edu.cn

Abstract

Background: The adoption rate of electronic health records (EHRs) in hospitals has become a main index to measure digitalization in medicine in each country.

Objective: This study summarizes and shares the experiences with EHR adoption in China and in the United States.

Methods: Using the 2007-2018 annual hospital survey data from the Chinese Health Information Management Association (CHIMA) and the 2008-2017 United States American Hospital Association Information Technology Supplement survey data, we compared the trends in EHR adoption rates in China and the United States. We then used the Bass model to fit these data and to analyze the modes of diffusion of EHRs in these 2 countries. Finally, using the 2007, 2010, and 2014 CHIMA and Healthcare Information and Management Systems Services survey data, we analyzed the major challenges faced by hospitals in China and the United States in developing health information technology.

Results: From 2007 to 2018, the average adoption rates of the sampled hospitals in China increased from 18.6% to 85.3%, compared to the increase from 9.4% to 96% in US hospitals from 2008 to 2017. The annual average adoption rates in Chinese and US hospitals were 6.1% and 9.6%, respectively. However, the annual average number of hospitals adopting EHRs was 1500 in China and 534 in the US, indicating that the former might require more effort. Both countries faced similar major challenges for hospital digitalization.

Conclusions: The adoption rates of hospital EHRs in China and the United States have both increased significantly in the past 10 years. The number of hospitals that adopted EHRs in China exceeded 16,000, which was 3.3 times that of the 4814 nonfederal US hospitals. This faster adoption outcome may have been a benefit of top-level design and government-led policies, particularly the inclusion of EHR adoption as an important indicator for performance evaluation and the appointment of public hospitals.

(J Med Internet Res 2021;23(2):e24813) doi:10.2196/24813

KEYWORDS

medical informatics; health information technologies; electronic health records; hospitals; Sino-American

Introduction

Electronic health records (EHRs) are the most important component of health information technology (HIT), and their adoption rate in hospitals indicates a country's level of digitalization in medicine. In the United States, EHRs enable the electronic documentation of providers' notes, electronic viewing of laboratory and radiology results, and electronic prescribing [1]. The 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act allocated approximately US $3 billion to accelerate the meaningful use of EHRs in US hospitals. Ultimately, the adoption rates in nonfederal hospitals increased from 9.4% in 2008 to 96% in 2017 [2,3]. In Asia, the EHR adoption rate increased from 15.1% in 2010 to 58.1% in 2015 in Korean hospitals [4] and from 21% in 2008 to 53% in 2014 in Japanese public hospitals [5]. In Europe, the usage of EHRs in German hospitals increased from 39.9% in 2007 to 68.4% in 2017 [6]. In China, the "Technical Specifications for Hospital Information Platforms based on EMRs" issued by the National Health Commission in 2015 defines electronic medical records (EMRs, corresponding to hospital EHRs) as complete and detailed clinical information resources that are created, stored, and used electronically by medical institutions and are generated and recorded for citizens in all visits to medical institutions [7]. Since 2015, the Chinese central government has invested over US $3.5 billion in HIT and EHRs and has issued 31 national policies and 134 technical standards covering all aspects of medical care digitalization and the construction of a digital medical security system. Thus, in China, EMRs are legal records created in hospitals and outpatient environments that constitute the data source of EHRs [8]. In the United States, the Promoting Interoperability Programs, led by the Centers for Medicare & Medicaid Services (CMS), do not specifically distinguish between EHRs and EMRs. In this study, the term EHRs refers specifically to the definitions provided by the CMS and China's National Health Commission.

Funding, policy, social organizations, and other factors, which can all greatly challenge any government, affect in-hospital EHR adoption. The most important factors associated with EHR adoption rates in hospitals are policy support and national standards. In the United States, relevant policies and standards include the HITECH Act [9], CMS Meaningful Use programs [10], and Promoting Interoperability Programs [8]. In China, they include the "46312" strategy [11], EMR Grading Evaluation Standards [12], and Hospital Intelligence Service Grading Evaluation Standards [13]. The second greatest factor affecting EHR promotion in both countries is insufficient financial support for digitalization in medicine [14]. Finally, another main issue is the large gap between the expectations of EHRs from clinical medical staff and their actual clinical performance.

As the world's largest country in terms of both population and number of hospitals, China has a unique medical system [15], with particular challenges affecting in-hospital EHR adoption. Therefore, the progress and difficulties in EHR adoption in Chinese hospitals are an important reference for other countries. First, through consecutive survey data analysis research, the EHR adoption in Chinese hospitals from 2007 to 2018 and the challenges of HIT innovation were summarized, based on the Chinese Health Information Management Association (CHIMA) Annual Survey—the longest and most authoritative national HIT industry survey in mainland China. Second, with the Bass model, we horizontally compared the EHR adoption rates of China and the United States from 2008 to 2017 and analyzed the challenges faced by the hospitals of these countries based on data taken from the Healthcare Information and Management Systems Services (HIMSS) Annual Surveys of 2007, 2010, and 2014. This study provides an overview and suggestions for further advancement of EHRs in hospitals in both countries, shares these experiences with other countries, and promotes global popularization of HIT.

Methods

EHR Definition and Function Reconciliation

Due to the differences in medical systems and traditions, a one-to-one mapping of the functions of EMRs in China and the United States is difficult. Nevertheless, we should clarify the definition and function of EHRs in these 2 countries so that the research results can reflect the closest comparable rates.

As for the United States, the EHR evaluation systems have, mainly, 2 aspects. First, for the governmental aspect, the Office of the National Coordinator for Health Information Technology (ONC) divided EHRs into "basic EHRs" (with or without clinical notes) and "comprehensive" EHRs in 2009. The former focuses on data collection and sharing and only needs to be implemented in one ward, while the latter stresses the clinical process based on the former and requires hospital coverage [2,16] (details in Multimedia Appendix 1). Since 2011, to facilitate the realization of a financial stimulus program, the CMS divided EHRs into 3 stages according to whether they are meaningfully used [17]. Each stage requires core objects and optional menu objects. Second, at the industry level, HIMSS Analytics developed an EMR adoption model (EMRAM) in 2005, including levels 0 to 7 based on "how many departments to use, standardization, sharing in hospital, decision support, sharing outside the hospital" [18].

As for China, the National Health Commission has been promoting the construction of EHRs with various policies and financial support since 2010 and issued the latest requirements on the definition and implementation timeline of EHRs in August 2018 [19]. In this requirement, EHR is divided into levels 0 to 8: levels 0 to 2 (low stage, focusing on the data collection function); levels 3 to 4 (medium stage, focusing on data sharing within or between departments and simple clinical decision making); and levels 5 to 8 (high stage, focusing on

clinical intelligent decision making, cross-hospital data sharing, and patient self-service; details in Multimedia Appendix 2).

There is no systematic comparative study of the evaluation systems of China and the United States. Our preliminary comparison study of EHRs in the top 2 tertiary hospitals in Beijing found that the Chinese EHR stage 4 hospitals can accomplish most (7 of 11) meaningfully used tasks in the United States [20]. However, the requirements for some specific functions of EHRs in the 2 countries are inconsistent, which complicates one-to-one matching of the 2 standards. Preliminarily, after comparing the common terms of the 2 standards, we think that Chinese EHR stages 3 and 4 roughly correspond to basic EHRs with notes and comprehensive EHRs, respectively. Unlike US EHR standards, Chinese EHR stage 1 requires Chinese hospitals to use the EHR for billing.

We did not use the data from HIMSS EMRAM, which was used in both countries as the research baseline, mainly because of the serious deviation of the sample distribution. Although about 74% of US hospitals passed HIMSS EMRAM stage 5 or above by the end of 2017, in China, the EMRAM is only a commercial trial project in a small number of hospitals. By June 2019, only 58 hospitals participated in the EMRAM evaluation and met or passed stage 6 [21].

Data Resources

Data on EHR adoption in Chinese hospitals were obtained from the CHIMA Annual Survey of Hospital Information Systems from 2007 to 2018 [22]. These are the only authoritative, national-level, long-term quantitative data of repeated surveys available on the EHR adoption rates in Chinese hospitals. Every March for a decade, the CHIMA surveyed the application of HIT in mainland China, covering 34 administrative regions. Survey respondents included general hospitals, teaching hospitals, specialty hospitals, traditional Chinese medicine hospitals, and integrated Chinese and Western medicine hospitals. In total, each survey was comprised of 9 parts. This study primarily used data from Parts I, IV, VI, and VIII, assessing respondents' basic information, information system application and adoption barriers, and data standardization.

Data on EHR adoption in US hospitals from 2008 to 2017 were obtained from data briefs by the ONC [2,3,16] and research by Jha et al [1,23-31]. Data on barriers faced in the information system implementation in US hospitals were obtained primarily from the HIMSS Annual Surveys in 2006, 2007, 2010, and 2014 [32-35].

Technology Diffusion Model and Bass Modeling

As one of our methods, Bass diffusion modeling was employed for the prediction and characterization of the progress in adoption of EHRs. Diffusion theory is an essential branch of communication theory [36]. The Bass model is widely used in the application and forecasting of new products and technologies [37,38], including many medical-related technologies [39-41]. The Bass model has 9 key assumptions [38,41], which mostly satisfy the scenarios of this study. For example, the market potential of a new product remains temporally constant; geographic boundaries of the social system are unchangeable throughout the diffusion.

Bass modeling has 2 important measures. First, the external influence coefficient, called the "innovation" effect and represented as the *p* coefficient, means the probability of using the product under the influence of public media or other external factors among users who have not used the product. Second, the internal influence coefficient, called the "imitation" effect and expressed as the *q* coefficient, depicts the probability of the same users using the product due to the influence of peers who have already used the product [42]. When *p* is high, the model indicates that the new technology has a rapid diffusion at the beginning of the propagation and that diffusion grows more weakly in the subsequent periods. When *q* is high, the model suggests that the new technology spreads slowly in the beginning, but it accelerates with further popularization and expansion. The Bass model is expressed as:

where *F(t)* is the portion of M adopted by time t, *p* is the coefficient of innovation, and *q* is the coefficient of imitation.

Data Analysis

We conducted statistical analyses and forecasts using linear optimization in Microsoft Excel for Mac 2011 (Microsoft Corporation, Redmond, WA). First, the analyses began with basic descriptive statistics regarding the respondents' basic information. Second, Bass diffusion modeling was employed to predict the progress of EHR adoption and analyze its characteristics. On one hand, we used the method of least squares to determine the optimal values of *q* and *p*. On the other hand, adjusted R^2 was used to evaluate the performance of the prediction model. The parameters of the Bass model were trained and estimated using SPSS 20 (IBM Corp, Armonk, NY).

Results

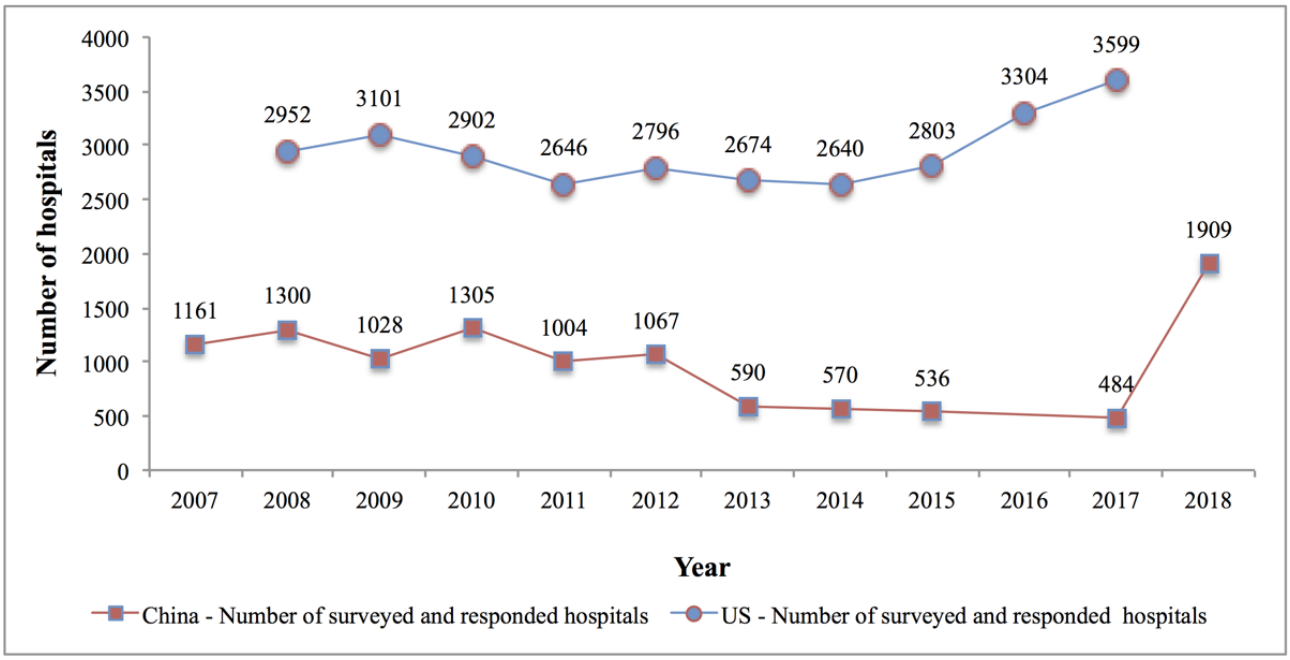
Descriptive Analysis

Scale and Coverage of the Surveys

Figure 1 illustrates the number of the 2007-2018 CHIMA Annual Survey respondents (covering over 80% of China's provinces, municipalities, and autonomous regions) and the number of survey respondents for the adoption of EHRs in US hospitals from 2008 to 2017. In China, all hospitals are classified by the government board into 3 classes: Level I hospitals (roughly equivalent to community-based health centers in the United States), Level II hospitals (county- and municipal-level, small health care facilities), and Level III hospitals (large, advanced general or specialty hospitals) [43]. In this study, hospitals were divided into 2 categories: Level III hospitals vs Level II or lower hospitals. For the definition of economically developed and underdeveloped areas in China, please refer to Multimedia Appendix 3. Data on the adoption of EHRs in US hospitals from 2008 to 2017 were obtained from the ONC data brief [2,3,16] and research by Jha et al [1,23-31], in which large hospitals were defined as those with ≥400 beds, while small and medium hospitals were those with 6-399 beds. Jha et al did not publish the number of surveyed and respondent hospitals in various subcategories in 2011 [28] and 2013 [26]. Since the ONC changed its statistical method after 2015, it only published

the overall EHR adoption rate of US hospitals but not the rates in various subcategories. Therefore, only the numbers of surveyed and respondent hospitals for 2016 and 2017 are included in Figure 1.

Figure 1. The number of respondents to the 2007-2018 Chinese Health Information Management Association (CHIMA) Annual Surveys of Hospital Information Systems and the surveys on the adoption of electronic health records (EHRs) in US hospitals from 2008 to 2017.

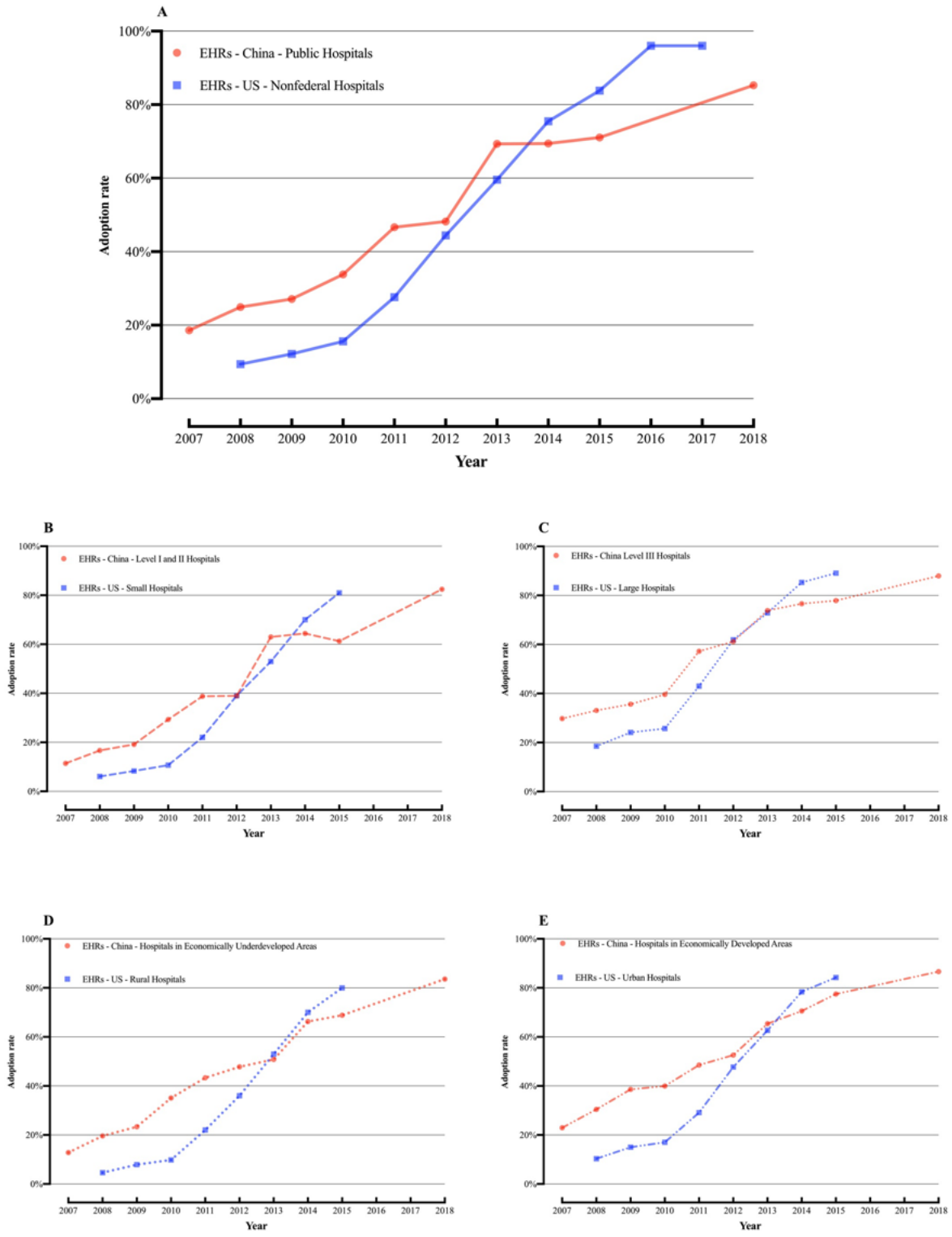


Detailed information about the scale of the Sino-American hospital (including the different hospital types) is provided in Multimedia Appendix 4.

Adoption of EHRs in Chinese and US Hospitals

Trends in EHR adoption in China and the United States were compared (Figure 2), which revealed 3 main characteristics. First, the EHR adoption rates in China were relatively high. Overall, the average EHR adoption rates of the sampled Chinese hospitals in 2018 (85.3%) were 1.5% higher than those of US hospitals in 2015 (83.8%), but lower than those of US hospitals in 2017 (96%). To note here, since the ONC changed its statistical method after 2015, it published only the overall EHR adoption rate of US hospitals but not the rates in various subcategories. Therefore, only the data for 2016 and 2017 are included in Figure 2A. Considering hospital scale, the adoption rates in Level II or lower Chinese hospitals (small-scale hospitals) were 1.5% higher than in small hospitals (fewer than 100 beds) in the United States in 2015 (the adoption rate of the former being 82.5% compared to the 81% of the latter). However, the average adoption rate of Level III hospitals in China (87.9%) was 1.2% lower than that of large US hospitals (89.1%). Considering regional economic development, the average adoption rate in Chinese hospitals in economically underdeveloped regions was 3.6% higher than in rural US hospitals—83.6% and 80%, respectively. The adoption rate in economically developed Chinese hospitals (86.6%) was 2.4% higher than that in urban US hospitals (84.2%).

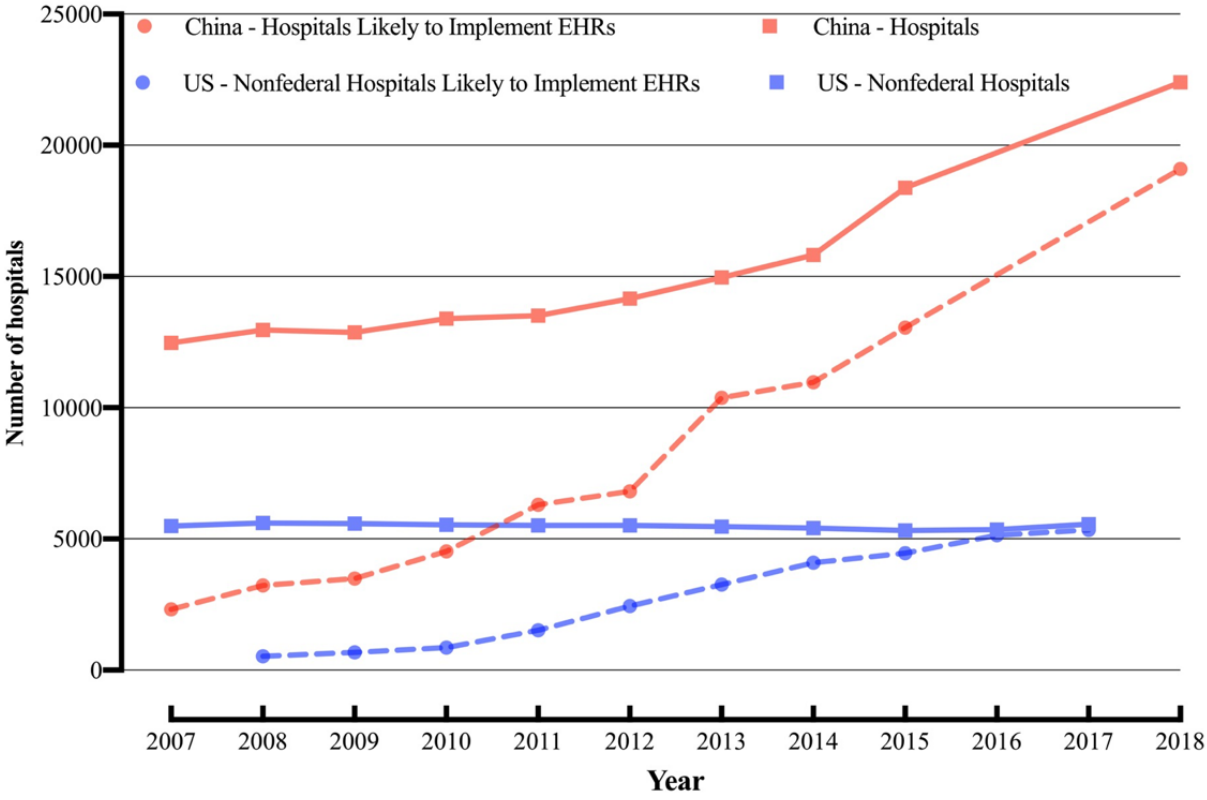
Figure 2. Trends in electronic health record (EHR) adoption rates in Chinese hospitals from 2007 to 2018 and nonfederal US hospitals from 2008 to 2017. (A) Overall adoption rate in China vs the United States and adoption rates in (B) small-scale hospitals, (C) large-scale hospitals, (D) hospitals in economically underdeveloped or rural areas, and (E) hospitals in economically developed or urban areas.



Because the overall number of hospitals in China exceeds the number of urban hospitals in the United States, the absolute number and challenges of Chinese hospitals adopting EHRs should be greater. The annual average number of hospitals adopting EHRs in China far exceeded the US average—1500 and 534, respectively. In 2007, China had 12,477 Level I-III hospitals [44] and an annual EHR adoption rate of 18.6%. According to sample projection, only 2322 Chinese hospitals used EHRs. In 2018, China had 22,396 Level I-III hospitals using EHRs [45], with an EHR adoption rate of 85.3% and a total of 19,094 hospitals. Thus, 16,772 hospitals in China adopted EHRs from 2007 to 2018—3.3 times the number of nonfederal hospitals adopting EHRs in the United States from

2008 to 2017 (4814), according to the projections based on the total number of nonfederal US hospitals (see Figure 3) [46].

Figure 3. Numbers of Chinese hospitals and those likely to implement electronic health records (EHRs) from 2007 to 2018 and the numbers of nonfederal US hospitals and those projected to implement EHRs from 2008 to 2017.



Difficulties With HIT Development in Chinese and US Hospitals

Figures 4 and 5 present the feedback from chief information officers (CIOs) on the barriers faced in the HIT application from the CHMIA and HIMSS surveys; as of 2015, the HIMSS Annual Survey no longer conducts a survey of hospital CIOs regarding the barriers to HIT application. Among Chinese hospitals, insufficient financial support and insufficient staff in the department were consistently identified as the first and second greatest obstacles, respectively. HIMSS Annual Survey data from the same years (2007, 2010, and 2014) show that US hospital CIOs also identified insufficient financial support and insufficient staff as their greatest challenges. This indicates a similarity in the main obstacles faced by China and the United States in hospital digitalization. In 2014, Chinese and US hospitals identified vendors' inability to deliver products and services to meet their demands as the third greatest obstacle. This may be because, with the increasing development of HIT in hospitals, hospital CIOs have become increasingly demanding with regard to the relevant software products.

Figure 4. Survey feedback on health information technology (HIT) development barriers faced by hospitals in China in 2007, 2010, and 2014. IT: information technology; ROI: return on investment.

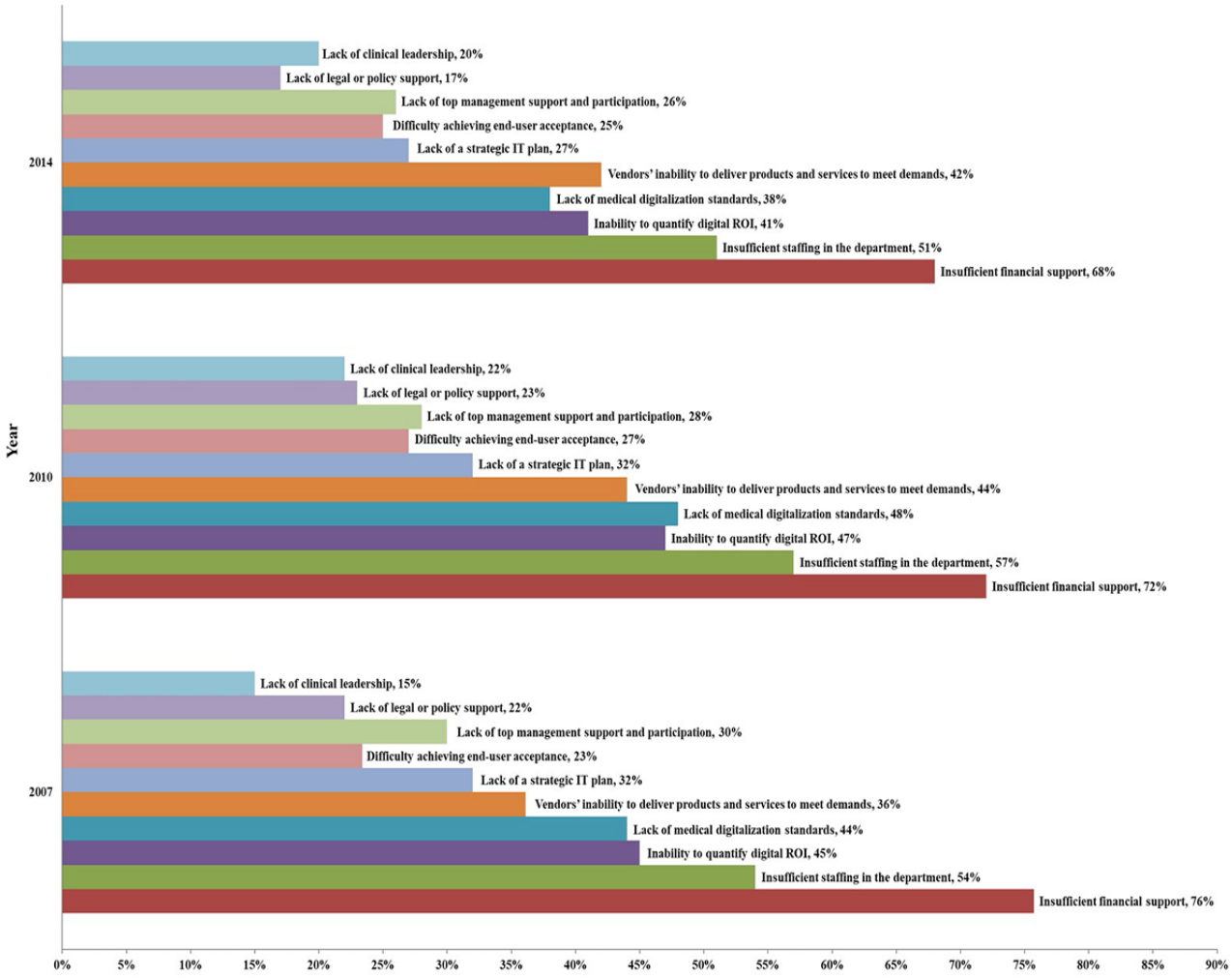
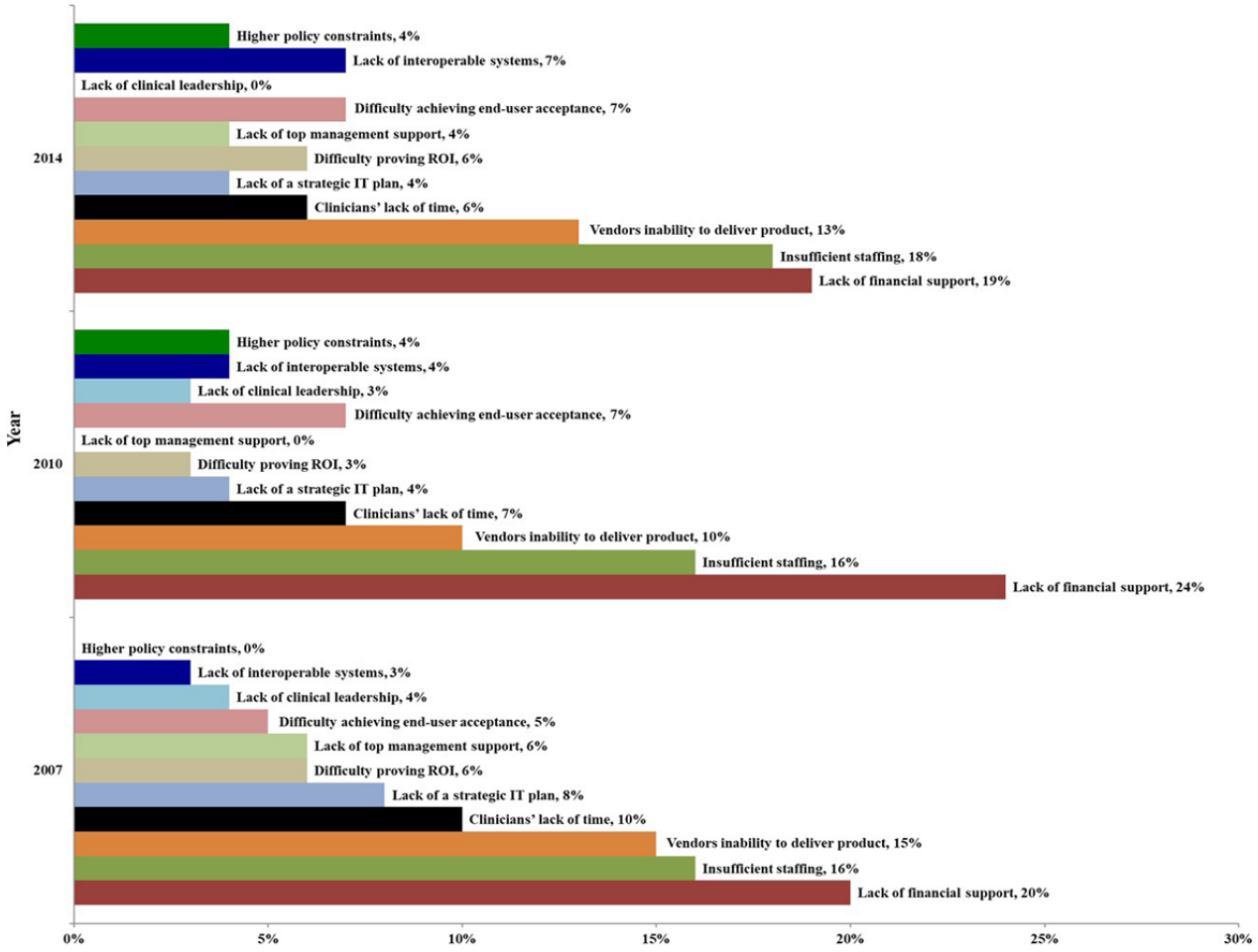


Figure 5. Survey feedback on health information technology (HIT) development barriers faced by hospitals in the United States in 2007, 2010, and 2014. IT: information technology; ROI: return on investment.



Bass Model Fitting and Prediction of EHR Adoption Rates in Chinese and US Hospitals

Considerable differences between Chinese and US hospitals in terms of the EHR technology diffusion modes were identified. Using Bass modeling and linear optimization, we estimated *p* and *q* coefficients based on the CHIMA data from 2007 to 2018 [22] and the surveys from 2008 to 2017 reported by the ONC [2,3,16] and Jha et al [1,23-31]. The parameter estimation results of the final model (Table 1) indicated that the Bass model fit the CHIMA dataset [22] and the ONC [2,3,16] and Jha et al [1,23-31] datasets. The adjusted R^2 was >0.9 for all models except the EHRs-China-Level III Hospitals model. Generally, each model shows a smaller motivation coefficient ratio (*q*/*p*) for Chinese hospitals compared to US hospitals. The largest difference (285-fold) was observed between the EHRs-China-Hospitals in the Economically Developed Areas model and the EHRs-US-Urban Hospitals model, which are the models of the largest-scale hospitals in these countries. In contrast, the smallest gap (14.8-fold) was observed between the EHRs-China-Surveyed Level I and II Hospitals model and the EHRs-US-Small Hospitals model, which are the models of these countries' smallest-scale hospitals. Moreover, the internal *q* of US hospitals was significantly larger than that of Chinese hospitals. The largest difference (57-fold) was observed between the EHRs-China-Hospitals in the Economically Developed Areas model and the EHRs-US-Urban Hospitals model, while the smallest (3.6-fold) was found between the EHRs-China-Level I and II hospitals model and the EHRs-US-Small Hospitals model.

Table 1. Bass model parameters for the prevalence of electronic health records (EHRs) in Chinese and US hospitals, based on Chinese Health Information Management Association (CHIMA) data from 2007 to 2018 [22] and survey data reported by the Office of the National Coordinator for Health Information Technology (ONC) [2,3,16] and Jha et al [1,23-31] from 2008 to 2017.

Model	Model parameters			
	p [a]	q [b]	q/p [c]	Adjusted R^2
EHRs-China	0.10	0.11	1.04	0.93
EHRs-China-Level I and II Hospitals	0.07	0.17	2.66	0.94
EHRs-China-Level III Hospitals	0.17	0.01	0.06	0.88
EHRs-China-Hospitals in Economically Underdeveloped Areas	0.08	0.13	1.51	0.98
EHRs-China-Hospitals in Economically Developed Areas	0.15	0.01	0.07	0.90
EHRs-US-Nonfederal Hospitals	0.02	0.58	24.33	0.97
EHRs-US-Small Hospitals	0.02	0.63	39.44	0.99
EHRs-US-Large Hospitals	0.07	0.45	6.64	0.95
EHRs-US-Rural Hospitals	0.01	0.65	46.29	0.99
EHRs-US-Urban Hospitals	0.03	0.57	19.66	0.98

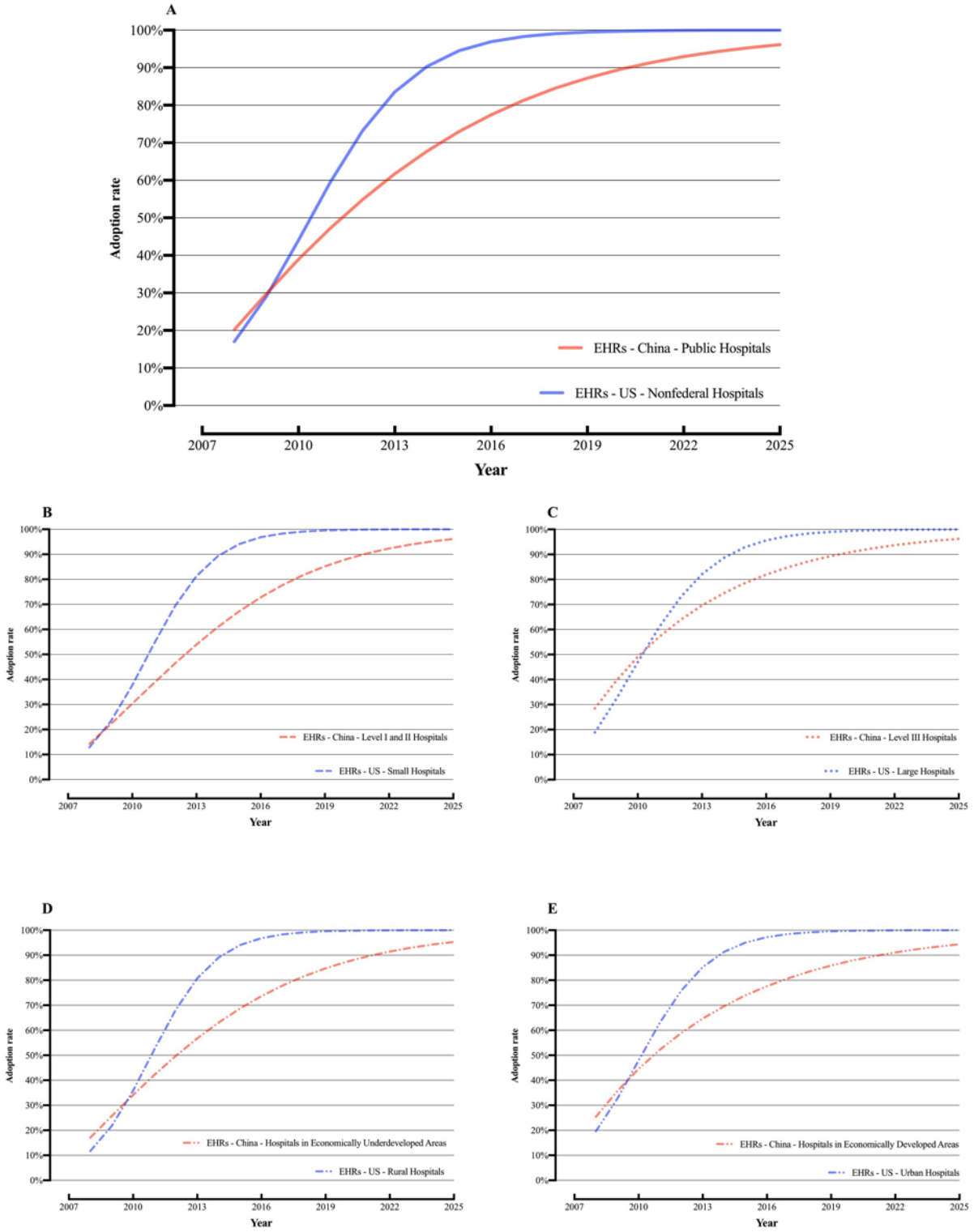
[a]p: external motivation coefficient.

[b]q: internal motivation coefficient.

[c]q/p: motivation coefficient ratio.

The differing diffusion patterns of EHRs in Chinese and US hospitals led to the differences in the patterns of the diffusion dynamics curves. By assuming that there will be no major policy adjustments or technological advancements in the future, we used the Bass model to fit and predict future EHR adoption in Chinese and US hospitals from 2019 to 2025, both in the overall scale and according to hospital scale and location (Figure 6). An Annual Survey was not conducted in 2016 due to changes in CHIMA's leadership. In 2017, the survey data from the software portion of the CHIMA Annual Survey deviated greatly; CHIMA does not recommend use of these data. Since the ONC changed its statistical method after 2015, it published only the overall EHR adoption rate of US hospitals but not the rates of various subcategories. Therefore, in Figures 6B, 6C, 6D, and 6E, the EHR adoption rates of various types of hospitals in the United States are predicted using the data from 2008 to 2015. The diffusion dynamics curve for EHRs in US hospitals forms a classic *S*-shape with a fast growth rate (p=0.03±0.02, q=0.58±0.07)—larger than that of Chinese hospitals (p=0.11±0.04, q=0.08±0.07).

Figure 6. Predicted electronic health record (EHR) adoption up to 2025 in Chinese hospitals (based on Chinese Health Information Management Association Annual Survey data from 2007 to 2018) and in US hospitals (based on annual survey data from 2008 to 2017 reported by the Office of the National Coordinator for Health Information Technology and Jha et al [1,23-31]). (A) Overall predictions and predictions for (B) small-scale hospitals, (C) large-scale hospitals, (D) hospitals in economically underdeveloped or rural areas, and (E) hospitals in economically developed or urban areas.



Discussion

Based on the 2007-2018 CHIMA Annual Surveys, we examined the progress and modes of EHR technology diffusion in sampled Chinese hospitals nationwide, identified major difficulties in HIT innovation, and compared them with US hospitals.

Principal Findings

From the perspective of EHR implementation in Chinese hospitals, Chinese hospitals demonstrated differences in EHR adoption and growth rates according to scale and location. Among the sampled hospitals in China, the adoption rates in small hospitals (Level II or lower) and in hospitals from

economically underdeveloped areas were below average. However, the growth rate of EHR adoption in these disadvantaged hospitals surpassed that of advantaged hospitals, as shown by the considerably higher slope in the Bass curves in Figure 6. This phenomenon is linked to national conditions, the medical system, and the financial support policy for HIT in hospitals in China. First, China has a vast territory that varies greatly from region to region. Although a large number of Level II or lower hospitals has been set up to provide basic medical services for local residents, high-quality medical resources are concentrated in a few Level III hospitals. Second, because China has not established a graded hierarchical medical system, patients are more inclined to congregate in Level III hospitals, leading to significantly higher workloads for doctors and correspondingly higher economic benefits [47,48]. As a result, many hospitals that are smaller or located in underdeveloped areas lack funds, resources, and motivation to build and maintain EHRs. Fortunately, the government has recognized this problem. Policies and funds should favor Level II or lower hospitals or those in underdeveloped areas, whereas Level III hospitals or those in developed areas should mostly be guided by policies and required to generate their own funding. Furthermore, the allowing of disadvantaged hospitals that have implemented EHRs to join medical institution alliances based on regional HIT has retained more patients in local hospitals instead of them seeking care in Level III hospitals, which also works to increase disadvantaged hospitals' income [11]. With the rapid development and wide application of wearable device technology [49,50], more real-time health data can be included in EHRs, which will further promote this trend.

From the perspective of the comparison of the EHR adoption by Chinese hospitals and US hospitals, although both the Chinese and US governments have implemented policy guidelines and financial incentives to promote EHR adoption, the patterns of EHR diffusion between the 2 countries differ considerably. The graph shapes in Figure 2 show this difference. The US trend is more *S*-shaped and more typical of a market-driven diffusion pattern, while the Chinese trend is more linear and more like a top-down, policy-driven pattern. EHR adoption in Chinese hospitals follows the innovator mode (motion coefficient ratio q/p is only 0.06 to 2.66), indicating that hospitals began to use EHRs in the initial stages due to the influence of external administrative forces [51]. This is because, in China, most (about 71.1%) of the secondary and higher hospitals are funded and managed by the government, and the number of beds in public hospitals is 3 times higher (about 76%) than that of private hospitals. In 2010, the government began to invest considerable resources and funds into EHRs and issued relevant policies to guide and support their use. However, the HIT support strategy at the time did not provide a detailed, clear, and measurable meaning of EHRs in Chinese hospitals nor establish any quantitative rewards, penalties, or standards for the use of EHRs by hospitals. This led to weak growth after the initial implementation of financial support. Reliance on hospital motivation to promote EHRs without sufficient external financial and policy incentives was proven to be unrealistic and unsuccessful [4,5,52,53]. In contrast, the EHR adoption rate in US hospitals grew very slowly from 2008 to 2010, perhaps due to the adoption of the US Health Insurance Portability and Accountability Act [54] in 1996—a comprehensive personal electronic health information privacy and security protection law—and the fact that most (about 80%) of the hospitals are private, which should be promoted by economic interests. Since 2010, following the implementation of meaningful use programs with clear quantitative requirements for EHRs, the EHR adoption rate increased significantly, and most US hospitals started to use certified EHRs by 2017. In sharp contrast to the low effectiveness of the expansive HIT development strategy in China, the HITECH Act was a major driving force behind this progress [55,56]. The US government provides financial incentives to US hospitals that implement EHRs and meet meaningful use phased standards and imposes financial penalties on those that do not [30]. We believe that the financial support and policy guidance of this "carrot and stick" model is also one of the most important American experiences.

The most significant morphological difference between the hospitals' EHR diffusion curves of China and the United States is that the motion coefficient ratio q/p value of the US curve is much larger. On one hand, the q value of the Chinese curve is smaller, and the P value is larger. This may be because the rapid spread of EHRs in China is caused by external policy stimuli. The Chinese government takes the HIT system represented by EHRs widely implemented in hospitals as a kind of technological innovation guided by the government and considers HIT as a technical tool to promote regional medical consortium [57]. As of 2015, according to our previous research results based on the same survey data, about 57.2% of the investigated hospitals have joined the regional medical consortium, 81.9% of these hospitals that have joined the medical consortium support the interconnection of electronic data, and the gap between HIT systems of different levels of medical institutions in the medical consortium is gradually narrowing [11]. On the other hand, the relatively large q value of the US curve may be interpreted in relation to 2 aspects. First, imitating the words or power of peers or industry leaders may influence American doctors to use similar EHRs as a tool for recording and exchanging health information [58]. Second, American doctors may have a strong willingness to upgrade information technology [59].

The comparison of the effects and outcomes of EHR implementation shows that the Chinese government has done more work to improve the implementation quantity and quality, as well as the relevant strategies used; has made unique contributions; and, thus, has had more achievements. This comparison can be made from 3 perspectives.

First, as for the implementation quantity, the EHR implementation rate in China in 2018 (85.3%) is equivalent to that in the United States in 2015 (83.8%) but is lower than that in the United States in 2017 (96%). However, since the base number of the former surpasses the latter (number of Chinese hospitals in 2018 was 22,396, compared to the number of nonfederal US hospitals in 2017, which was 5564) [46], the number of hospitals adopting EHRs in Chinese hospitals is approximately 3.3 times that in the United States—16,772 and 4818, respectively. Moreover, the annual growth of the former (1500) is about 2.8 times that of the latter (534). As of 2018, although China's population (1.4 billion) is 4.28 times that of

the United States (327 million), given that the Chinese gross domestic product (GDP) is only 67.8% that of the latter (the Chinese GDP being US $13.89 trillion, compared to the US $20.5 trillion GDP in the United States), the per capita GDP of China (US $9900) is only 15.8% that of the latter (US $62,500) and is below the world average (US $11,300). It also reflects that the former has made great progress in promoting EHRs in a short time (11 years), under the promotion of a huge subjective initiative.

Second, as for the implementation quality, the China Health Commission has conducted many top-level design policies. First, the connotation of EHRs is clearly defined, and EHR adoption is divided into Levels 0 to 8. Second, through administrative instructions, different deadlines are set for hospitals at different levels. For example, by the end of 2020, all Level III and Level II hospitals must use at least Level IV and Level III EHRs, respectively. Namely, by the end of 2020 [60], 11,565 secondary and tertiary hospitals in China, accounting for 52% of the country's 22,000 hospitals, must use at least Chinese stage 3 EHRs (roughly corresponding to basic EHRs with notes). This is 2.4 times the number of the 4818 nonfederal hospitals implementing EHRs in the United States in 2017 [2,16]. As of July 2020, 128 Chinese hospitals were tested and verified using EHRs that met the high-level (stages 5-7) standards—44 more than the same period of last year (84 hospitals)—of which 4 reached stage 7 (an increase of 2 hospitals) and 20 reached stage 6 (an increase of 15 hospitals) [61]. Moreover, the performance monitoring data of the Chinese government for public hospitals partially verify and support the prediction results of the BASS model. As of July 2020, the announcement on "the National Monitoring and Analysis of the Performance Appraisal of the National Tertiary Public Hospitals" released by the Chinese government in 2018 [62] shows that the participation rate of China's EMR level evaluation of tertiary hospitals was 94.58% by 2018, with an average stage of 2.72—a stage close to the level of basic EHR with clinical notes in the United States. Approximately 87% of tertiary hospitals reached EHR Level III or above, which is very close to the prediction result of the Bass model (87.2%). We believe that as of the end of 2020, China's tertiary hospitals were likely to achieve stage 4 EMRs (namely, comprehensive EHRs). The Chinese government has also released information, which was published near the end of 2020, on the results of the performance appraisal of hospitals, including the progress of the implementation of EMRs in public hospitals below the third level.

Third, as for the implementation strategy, evaluation is emphasized with the principle "promote construction with evaluation, promote improvement with evaluation." The Chinese government has adopted different direct capital investments and indirect policy guidance strategies for hospitals of different scales. For the Level III large hospitals, which are responsible for over 46% of outpatients in China, the government focused on policy guidance, released many guidelines [63] and management and normative documents from 2010 to 2019 to promote hospital digitization with EHRs as the core, and encouraged private capital investments [64]. Additionally, for small hospitals at a level lower than Level III, a strategy of direct finance and indirect guidance was adopted to gradually promote EHR implementation. Furthermore, in 2019, the State Council of China stipulated that the construction of EHRs was one major indicator for hospital-level assessment and appointment of public hospital presidents [64]. For example, in the 3-year national hospital evaluation, the EHRs used by tertiary hospitals must meet Chinese stage 4; otherwise, the hospitals will be downgraded, which will greatly affect the reputation and economic income of the hospitals.

Limitations

The data used here were collected from (1) repeated measurements of EHR constructions in the same batch of US hospitals affiliated with the ONC and American Hospital Association (2007-2017) and (2) repeated investigations through self-report questionnaires (2007-2015, 2018) of EHR construction from Chinese hospitals participating in annual conferences organized by the CHIMA. The latter was not independently verified. Therefore, such analysis might be affected by several potential confounding factors of data bias. First, due to the limitation of the CHIMA survey data, there may be limitations to the classification of EHRs in hospitals in China. However, the implementation rate of each classification of hospitals is only an added reference index. Moreover, there are no such hospitals in the United States, so this classification is not used mainly for the comparison of the same hospitals in China and the United States. Second, we did not use multivariate models to assess the independence among different factors (eg, grades, types, economic levels, or locations of hospitals). Third, the cumulative proportions of some repeated questionnaire data from CHIMA during 2007-2015 slightly declined. We think one explanation may be that throughout the repeated surveys, the sampling differences of hospital samples led to differences in the investigated data. Although we limit our deductions to our own samples, our analyses are valuable in that these data are the only available quantitative data concerning the trend in HIT development in China over a time span of 10 years and collected by the Chinese state-level academy in this field. These are the only authoritative, national-level, long-term quantitative data available on the rate of adoption of EHRs in Chinese hospitals. Therefore, these are the best available data that can reflect the status of EHR use in Chinese hospitals. Furthermore, due to the differences in the economic, cultural, and health systems between China and the United States, there are also some differences in the functional definition of necessary components of hospital EHRs. Therefore, we mainly analyzed the overall time trend of EHR implementation in hospitals of the 2 countries, and the horizontal comparison is only for an approximate reference.

Conclusion

Over the last decade, the Chinese government has identified HIT development, represented by hospital EHRs, as an important technical focus and starting point to support medical reform. According to the CHIMA Annual Surveys, the average EHR adoption rate in sampled hospitals in China increased by 3.6 times from 2007 to 2018, peaking at 85.3%, which exceeds that of 83.8% in US hospitals in 2015 but is lower than the 96% recorded in 2017. The difference in the EHR technology

diffusion curves of China and the United States based on the Bass model is very likely due to the differences in the EHR promotion, implementation, and management policies, as well as the medical system, of the 2 countries. The former is mainly stimulated by external policies, while the latter is initiated by their own technological upgrading needs. The Chinese government has begun to amend relevant policies, gradually implementing both financial support and policy guidance measures and adding the assessment of secondary utilization based on precipitated data on EHRs and the use of various advanced functions. This action technically underlies several medical reform goals, such as improving clinical outcomes, user satisfaction, and interoperability. Various signs indicate that the Chinese government is gradually approaching and realizing its phase goals established in the second medical reform initiated in 2010, including the integration of medical resources, improvement of the popularization and quality of medical care, and the reduction of medical costs.

Acknowledgments

We thank Elsevier language editing. This work was supported by the National Natural Science Foundation of China (grant numbers 81771937, 81871455) and partly supported by PKU-Baidu Fund（project of intelligent auxiliary diagnosis using medical images）and Luzhou Science and Technology Bureau Fund (research project of constructing health big data platforms and service systems for combined medical and nursing institutions for elderly care).

Authors' Contributions

The work presented here was carried out in collaboration among all authors. JL conceived and designed the study. JL, XZ, TW, and YL performed the literature review and undertook data acquisition and data analysis. JL drafted the manuscript, and JL and JZ significantly revised the manuscript. ZZ, DS, BT, and JX supervised the review methodology and data interpretation and supplied valuable suggestions for improvement.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Electronic health record (EHR) functions used to define "basic without clinical notes," "basic with clinical notes," and "comprehensive" EHR systems.
[DOCX File , 21 KB - jmir_v23i2e24813_app1.docx]

Multimedia Appendix 2
Chinese evaluation and management of application level of EHRs system.
[DOCX File , 18 KB - jmir_v23i2e24813_app2.docx]

Multimedia Appendix 3
The definitions of economically developed and underdeveloped areas in China from CHIMA Annual Surveys of Hospital Information Systems.
[DOCX File , 16 KB - jmir_v23i2e24813_app3.docx]

Multimedia Appendix 4
Scale of the 2007-2018 CHIMA Annual Surveys of Hospital Information Systems and the 2008-2017 surveys on adoption of EHRs in U.S. hospitals.
[DOCX File , 19 KB - jmir_v23i2e24813_app4.docx]

References

1. Jha AK, Ferris TG, Donelan K, DesRoches C, Shields A, Rosenbaum S, et al. How common are electronic health records in the United States? A summary of the evidence. Health Aff (Millwood) 2006 Dec;25(6):w496-w507. [doi: 10.1377/hlthaff.25.w496] [Medline: 17035341]
2. Henry J, Pylypchuk Y, Searcy T, Patel V. Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015. Office of the National Coordinator for Health Information Technology. 2016 May. URL: https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php [accessed 2021-02-06]
3. Non-federal Acute Care Hospital Health IT Adoption and Use. Office of the National Coordinator for Health Information Technology. 2019 Nov. URL: https://healthdata.gov/dataset/non-federal-acute-care-hospital-health-it-adoption-and-use [accessed 2021-02-06]

4. Kim Y, Jung K, Park Y, Shin D, Cho S, Yoon D, et al. Rate of electronic health record adoption in South Korea: A nation-wide survey. Int J Med Inform 2017 May;101:100-107 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.02.009] [Medline: 28347440]
5. Kanakubo T, Kharrazi H. Comparing the Trends of Electronic Health Record Adoption Among Hospitals of the United States and Japan. J Med Syst 2019 Jun 11;43(7):224. [doi: 10.1007/s10916-019-1361-y] [Medline: 31187293]
6. Esdar M, Hüsers J, Weiß JP, Rauch J, Hübner U. Diffusion dynamics of electronic health records: A longitudinal observational study comparing data from hospitals in Germany and the United States. Int J Med Inform 2019 Nov;131:103952 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.103952] [Medline: 31557699]
7. Technical specification of hospital information platform based on electronic medical record. National Health Commission of the People's Republic of China. 2015. URL: http://www.nhc.gov.cn/wjw/s9497/201406/a2014514701f4e76b14f3446f6318937.shtml [accessed 2021-02-06]
8. Garets D, Davis M. Electronic medical records vs electronic health records: yes, there is a difference. American Academy of Ophthalmology. 2006 Jan 26. URL: https://www.aao.org/asset.axd?id=8e9b1f20-0ed6-4d2b-92f8-e28fbaf378ec&t=634962799822530000 [accessed 2021-02-06]
9. Gold M, McLaughlin C. Assessing HITECH Implementation and Lessons: 5 Years Later. Milbank Q 2016 Sep;94(3):654-687 [FREE Full text] [doi: 10.1111/1468-0009.12214] [Medline: 27620687]
10. Colicchio TK, Cimino JJ, Del Fiol G. Unintended Consequences of Nationwide Electronic Health Record Adoption: Challenges and Opportunities in the Post-Meaningful Use Era. J Med Internet Res 2019 Jun 03;21(6):e13313 [FREE Full text] [doi: 10.2196/13313] [Medline: 31162125]
11. Liang J, Zheng X, Chen Z, Dai S, Xu J, Ye H, et al. The experience and challenges of healthcare-reform-driven medical consortia and Regional Health Information Technologies in China: A longitudinal study. Int J Med Inform 2019 Nov;131:103954 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.103954] [Medline: 31513943]
12. Electronic Medical Record System Application Level Grading Evaluation Standard (Trial). National Health Commission of the People's Republic of China. 2018 Dec 3. URL: http://www.nhc.gov.cn/yzygj/s7659/201812/3cae6834a65d48e9bfd783f3c7d54745.shtml [accessed 2019-09-01]
13. Hospital Smart Service Rating Evaluation Standard System (Trial). National Health Commission of the People's Republic of China. 2019 May 05. URL: http://wjw.fy.gov.cn/content/detail/5cd372d97f8b9a55448b457f.html [accessed 2021-02-06]
14. Liang J, Li Y, Zhang Z, Shen D, Xu J, Yu G, et al. Evaluating the Applications of Health Information Technologies in China During the Past 11 Years: Consecutive Survey Data Analysis. JMIR Med Inform 2020 Feb 10;8(2):e17006 [FREE Full text] [doi: 10.2196/17006] [Medline: 32039815]
15. Feng Z. Health Care For 1.4 Billion: China's Healthcare System And Reform. Health Affairs 2018 Feb;37(2):333-334. [doi: 10.1377/hlthaff.2017.1526]
16. Charles D, Gabriel M, Searcy T. Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2014. Office of the National Coordinator for Health Information Technology. 2015 Apr. URL: https://www.healthit.gov/sites/default/files/data-brief/2014HospitalAdoptionDataBrief.pdf [accessed 2021-02-06]
17. What is meaningful use? The Office of the National Coordinator for Health Information Technology. 2013 Jun 1. URL: https://www.healthit.gov/faq/what-meaningful-use [accessed 2020-06-11]
18. Electronic Medical Record Adoption Model. HIMSS Analytics. 2015. URL: https://www.himssanalytics.org/asia-pacific/electronic-medical-record-adoption-model [accessed 2021-02-08]
19. Notice on Further Promoting the digitalization Construction of Medical Institutions with Electronic Medical Records as the Core. National Health Commission of the People's Republic of China. 2018 Aug 28. URL: http://www.nhc.gov.cn/yzygj/s7659/201808/a924c197326440cdaaa0e563f5b111c2.shtml [accessed 2020-06-01]
20. Xu L, Wen D, Zhang X, Lei J. Assessing and comparing the usability of Chinese EHRs used in two Peking University hospitals to EHRs used in the US: A method of RUA. Int J Med Inform 2016 May;89:32-42. [doi: 10.1016/j.ijmedinf.2016.02.008] [Medline: 26980357]
21. Healthcare Information and Management Systems Society. Public hospitals are not allowed to participate in HIMSS ratings. MedSci/Policy and Humanities. 2019. URL: https://www.medsci.cn/article/show_article.do?id=b49f1e528663 [accessed 2020-06-21]
22. Luo S, Zhang K, Li B. Medical informatics in China: healthcare IT trends, academic and research developments. Yearb Med Inform 2010:109-115. [Medline: 20938582]
23. Sheikh A, Jha A, Cresswell K, Greaves F, Bates DW. Adoption of electronic health records in UK hospitals: lessons from the USA. The Lancet 2014 Jul 05;384(9937):8-9. [doi: 10.1016/S0140-6736(14)61099-0] [Medline: 24998803]
24. Furukawa MF, King J, Patel V, Hsiao C, Adler-Milstein J, Jha AK. Despite substantial progress In EHR adoption, health information exchange and patient engagement remain low in office settings. Health Aff (Millwood) 2014 Sep;33(9):1672-1679. [doi: 10.1377/hlthaff.2014.0445] [Medline: 25104827]
25. Adler-Milstein J, DesRoches CM, Kralovec P, Foster G, Worzala C, Charles D, et al. Electronic Health Record Adoption In US Hospitals: Progress Continues, But Challenges Persist. Health Aff (Millwood) 2015 Dec;34(12):2174-2180. [doi: 10.1377/hlthaff.2015.0992] [Medline: 26561387]

26. Adler-Milstein J, DesRoches CM, Furukawa MF, Worzala C, Charles D, Kralovec P, et al. More than half of US hospitals have at least a basic EHR, but stage 2 criteria remain challenging for most. Health Aff (Millwood) 2014 Sep;33(9):1664-1671. [doi: 10.1377/hlthaff.2014.0453] [Medline: 25104826]
27. Ford EW, Menachemi N, Peterson LT, Huerta TR. Resistance is futile: but it is slowing the pace of EHR adoption nonetheless. J Am Med Inform Assoc 2009 May;16(3):274-281 [FREE Full text] [doi: 10.1197/jamia.M3042] [Medline: 19261931]
28. DesRoches CM, Worzala C, Joshi MS, Kralovec PD, Jha AK. Small, nonteaching, and rural hospitals continue to be slow in adopting electronic health record systems. Health Aff (Millwood) 2012 May;31(5):1092-1099 [FREE Full text] [doi: 10.1377/hlthaff.2012.0153] [Medline: 22535503]
29. Jha AK, DesRoches CM, Campbell EG, Donelan K, Rao SR, Ferris TG, et al. Use of electronic health records in U.S. hospitals. N Engl J Med 2009 Apr 16;360(16):1628-1638. [doi: 10.1056/NEJMsa0900592] [Medline: 19321858]
30. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide. J Am Med Inform Assoc 2017 Nov 01;24(6):1142-1148 [FREE Full text] [doi: 10.1093/jamia/ocx080] [Medline: 29016973]
31. Jha AK, Burke MF, DesRoches C, Joshi MS, Kralovec PD, Campbell EG, et al. Progress toward meaningful use: hospitals' adoption of electronic health records. Am J Manag Care 2011 Dec;17(12 Spec No):SP117-SP124 [FREE Full text] [Medline: 22216770]
32. Healthcare Information and Management Systems Society. 17th Annual 2006 HIMSS Leadership Survey, Healthcare CIO Final Report. HIMSS February 13. Report No 2006:1-16.
33. Healthcare Information and Management Systems Society. 18th Annual 2007 HIMSS Leadership Survey, Healthcare CIO Final Report. HIMSS April 10. Report No 2007:1-15.
34. Healthcare Information and Management Systems Society. 21th Annual 2010 HIMSS Leadership Survey, Healthcare CIO Final Report. HIMSS February 19. Report No 2010:1-18.
35. Healthcare Information and Management Systems Society. 25th Annual 2014 HIMSS Leadership Survey Healthcare CIO Final Report. HIMSS February 24. Report No 2014:1-38.
36. Bass FM. A New Product Growth for Model Consumer Durables. Management Science 2004 Dec;50(12_supplement):1825-1832. [doi: 10.1287/mnsc.1040.0264]
37. Van de Bulte C. Want to know how diffusion speed varies across countries and products? Try using a Bass model. PDMA visions 2002;26(4):12-15. [doi: 10.1002/9781444316568.wiem05030]
38. Sood A, James GM, Tellis GJ, Zhu J. Predicting the Path of Technological Innovation: SAW vs. Moore, Bass, Gompertz, and Kryder. Marketing Science 2012 Nov;31(6):964-979. [doi: 10.1287/mksc.1120.0739]
39. Kharrazi H, Gonzalez CP, Lowe KB, Huerta TR, Ford EW. Forecasting the Maturation of Electronic Health Record Functions Among US Hospitals: Retrospective Analysis and Predictive Model. J Med Internet Res 2018 Aug 07;20(8):e10458 [FREE Full text] [doi: 10.2196/10458] [Medline: 30087090]
40. Ford EW, Hesse BW, Huerta TR. Personal Health Record Use in the United States: Forecasting Future Adoption Levels. J Med Internet Res 2016 Mar 30;18(3):e73 [FREE Full text] [doi: 10.2196/jmir.4973] [Medline: 27030105]
41. Norton JA, Bass FM. A Diffusion Theory Model of Adoption and Substitution for Successive Generations of High-Technology Products. Management Science 1987 Sep;33(9):1069-1086. [doi: 10.1287/mnsc.33.9.1069]
42. Grimm SE, Stevens JW, Dixon S. Estimating Future Health Technology Diffusion Using Expert Beliefs Calibrated to an Established Diffusion Model. Value Health 2018 Aug;21(8):944-950 [FREE Full text] [doi: 10.1016/j.jval.2018.01.010] [Medline: 30098672]
43. Lei J, Guan P, Gao K, Lu X, Chen Y, Li Y, et al. Characteristics of health IT outage and suggested risk management strategies: an analysis of historical incident reports in China. Int J Med Inform 2014 Feb;83(2):122-130. [doi: 10.1016/j.ijmedinf.2013.10.006] [Medline: 24246272]
44. National Health Commission of the People's Republic of China. In: Chen Z, editor. 2010 China Health Statistics Yearbook. Beijing: Peking Union Medical College Press; 2010.
45. National Health Commission of the People's Republic of China. In: Ma X, editor. 2019 China Health Statistics Yearbook. Beijing: Peking Union Medical College Press; 2019.
46. Elflein J. Number of federal and non-feaderal hospitals in the US from 1957 to 2017. Statista. 2020 Jan. URL: https://www.statista.com/statistics/185852/number-of-federal-and-nonfederal-hospitals-in-the-us-since-2001/ [accessed 2020-02-05]
47. Fu Y, Schwebel D, Hu G. Physicians' Workloads in China: 1998 2016. Int J Environ Res Public Health 2018 Aug 03;15(8):1649 [FREE Full text] [doi: 10.3390/ijerph15081649] [Medline: 30081497]
48. Wen J, Cheng Y, Hu X, Yuan P, Hao T, Shi Y. Workload, burnout, and medical mistakes among physicians in China: A cross-sectional study. Biosci Trends 2016 Feb;10(1):27-33 [FREE Full text] [doi: 10.5582/bst.2015.01175] [Medline: 26961213]
49. Wen D, Zhang X, Liu X, Lei J. Evaluating the Consistency of Current Mainstream Wearable Devices in Health Monitoring: A Comparison Under Free-Living Conditions. J Med Internet Res 2017 Mar 07;19(3):e68 [FREE Full text] [doi: 10.2196/jmir.6874] [Medline: 28270382]

50. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e94 [FREE Full text] [doi: 10.2196/mhealth.9754] [Medline: 29650506]
51. Rogers EM. Diffusion of Innovations, 5th Edition. NewYork, NY: Free Press; 2003.
52. Yoon D, Chang B, Kang S, Bae H, Park R. Adoption of electronic health records in Korean tertiary teaching and general hospitals. Int J Med Inform 2012 Mar;81(3):196-203 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.12.002] [Medline: 22206619]
53. Yoshida Y, Imai T, Ohe K. The trends in EMR and CPOE adoption in Japan under the national strategy. Int J Med Inform 2013 Oct;82(10):1004-1011 [FREE Full text] [doi: 10.1016/j.ijmedinf.2013.07.004] [Medline: 23932755]
54. Bendix J. HIPAA: how to protect yourself and your practice. Med Econ 2013 Aug 10;90(15):38, 40-38, 41. [Medline: 24730106]
55. Cohen MF. Impact of the HITECH financial incentives on EHR adoption in small, physician-owned practices. Int J Med Inform 2016 Oct;94:143-154. [doi: 10.1016/j.ijmedinf.2016.06.017] [Medline: 27573322]
56. Mennemeyer ST, Menachemi N, Rahurkar S, Ford EW. Impact of the HITECH Act on physicians' adoption of electronic health records. J Am Med Inform Assoc 2016 Mar;23(2):375-379 [FREE Full text] [doi: 10.1093/jamia/ocv103] [Medline: 26228764]
57. Wang Z. Data integration of electronic medical record under administrative decentralization of medical insurance and healthcare in China: a case study. Isr J Health Policy Res 2019 Apr 01;8(1):24 [FREE Full text] [doi: 10.1186/s13584-019-0293-9] [Medline: 30929644]
58. Yaraghi N, Du AY, Sharman R, Gopal RD, Ramesh R, Singh R, et al. Professional and geographical network effects on healthcare information exchange growth: does proximity really matter? J Am Med Inform Assoc 2014;21(4):671-678 [FREE Full text] [doi: 10.1136/amiajnl-2012-001293] [Medline: 24287171]
59. Whitney E. Sharing Patient Records Is Still A Digital Dilemma For Doctors. Morning Edition. 2015 Mar 06. URL: http://www.npr.org/blogs/health/2015/03/06/388999602/sharing-patientrecords-is-still-a-digital-dilemma-for-doctors [accessed 2021-02-06]
60. Notice on the issuance of the administrative measures for grading evaluation of the application level of electronic medical records system (trial implementation) and evaluation standards (trial implementation). National Health Commission of the People's Republic of China. 2018 Dec 19. URL: http://www.gov.cn/xinwen/2018-12/09/content_5347261.htm [accessed 2020-06-22]
61. Full analysis of 128 high-level evaluation hospitals with electronic medical records. CN-Healthcare. 2020 Jul 17. URL: https://www.cn-healthcare.com/article/20200717/content-539591.html [accessed 2020-07-31]
62. Announcement on the National Monitoring and Analysis of the Performance Appraisal of the National Tertiary Public Hospitals in 2018. National Health Commission of the People's Republic of China. 2020 Jul 01. URL: http://www.nhc.gov.cn/yzygj/s3593g/202006/863717ce64af4372a737048cf500eb3d.shtml [accessed 2021-02-06]
63. Fu Y, Schwebel D, Hu G. Physicians' Workloads in China: 1998 2016. Int J Environ Res Public Health 2018 Aug 03;15(8):1649 [FREE Full text] [doi: 10.3390/ijerph15081649] [Medline: 30081497]
64. Ting S, Haiyi L, Wei Z. Enlightenment from the Development of Chinese and American Electronic Medical Record System in the Past Decade. China Digital Medicine 2019;14(11):10-13.

Abbreviations

CHIMA: Chinese Hospital Information Management Association
CIO: chief information officer
CMS: Centers for Medicare & Medicaid Services
EHR: electronic health record
EMR: electronic medical record
EMRAM: EMR adoption model
GDP: gross domestic product
HIT: health information technology
HITECH: Health Information Technology for Economic and Clinical Health
ONC: Office of the National Coordinator for Health Information Technology

Edited by G Eysenbach; submitted 07.10.20; peer-reviewed by H Wei, T Hao, W Fuzhi, J Cimino; comments to author 23.10.20; revised version received 29.11.20; accepted 21.01.21; published 18.02.21.

Please cite as:
Liang J, Li Y, Zhang Z, Shen D, Xu J, Zheng X, Wang T, Tang B, Lei J, Zhang J
Adoption of Electronic Health Records (EHRs) in China During the Past 10 Years: Consecutive Survey Data Analysis and Comparison of Sino-American Challenges and Experiences
J Med Internet Res 2021;23(2):e24813
URL: http://www.jmir.org/2021/2/e24813/
doi:10.2196/24813
PMID:33599615

Original Paper

Patterns and Influencing Factors of eHealth Tools Adoption Among Medicaid and Non-Medicaid Populations From the Health Information National Trends Survey (HINTS) 2017-2019: Questionnaire Study

Xin Yang[1], DPhil; Ning Yang[2], MSc; Dwight Lewis[1,3], DPhil; Jason Parton[1,2], DPhil; Matthew Hudnall[1,2], DPhil

[1]Institute of Data and Analytics, The University of Alabama, Tuscaloosa, AL, United States

[2]Department of Information Systems, Statistics, and Management Science, Culverhouse College of Business, The University of Alabama, Tuscaloosa, AL, United States

[3]Department of Management, Culverhouse College of Business, The University of Alabama, Tuscaloosa, AL, United States

Corresponding Author:
Xin Yang, DPhil
Institute of Data and Analytics
The University of Alabama
250 Bidgood Hall
Tuscaloosa, AL, 35406
United States
Phone: 1 2053483267
Email: xyang15@cba.ua.edu

Abstract

Background: Evidence suggests that eHealth tools adoption is associated with better health outcomes among various populations. The patterns and factors influencing eHealth adoption among the US Medicaid population remain obscure.

Objective: The objective of this study is to explore patterns of eHealth tools adoption among the Medicaid population and examine factors associated with eHealth adoption.

Methods: Data from the Health Information National Trends Survey from 2017 to 2019 were used to estimate the patterns of eHealth tools adoption among Medicaid and non-Medicaid populations. The effects of Medicaid insurance status and other influencing factors were assessed with logistic regression models.

Results: Compared with the non-Medicaid population, the Medicaid beneficiaries had significantly lower eHealth tools adoption rates for health information management (11.2% to 17.5% less) and mobile health for self-regulation (0.8% to 9.7% less). Conversely, the Medicaid population had significantly higher adoption rates for using social media for health information than their counterpart (8% higher in 2018, *P*=.01; 10.1% higher in 2019, *P*=.01). Internet access diversity, education, and cardiovascular diseases were positively associated with health information management and mobile health for self-regulation among the Medicaid population. Internet access diversity is the only factor significantly associated with social media adoption for acquisition of health information (OR 1.98, 95% CI 1.26-3.11).

Conclusions: Our results suggest digital disparities in eHealth tools adoption between the Medicaid and non-Medicaid populations. Future research should investigate behavioral correlates and develop interventions to improve eHealth adoption and use among underserved communities.

(J Med Internet Res 2021;23(2):e25809) doi:10.2196/25809

KEYWORDS

Medicaid program; eHealth; internet access; digital divide; health information technology

Introduction

Though progress has been made with improving the “digital divide” in the United States, disparities in internet access remain among underserved populations such as the elderly and individuals in lower socioeconomic classifications [1]. This digital divide can have profound implications on public health because internet access is hypothesized as a “super-determinant”

of health [2]. One primary way by which the internet may improve health outcomes is through the adoption of electronic health (eHealth). eHealth is the application and integration of information and communication technologies (ICT) to improve patients' health care, enable more reliable connections between patients and health providers, and reduce health disparities [3,4].

Standard eHealth tools used by consumers include the utilization of electronic health records (EHRs), health-related social media channels to seek or share health information, and online patient-provider communication (OPPC) channels such as online portals, email, and social media [5]. Recent advancements in information technology (IT) expanded the adoption of eHealth tools to smart devices such as mobile health tools (eg, smartphone apps, wearable devices such as Fitbit, and smartwatches) [6] and virtual assistants (eg, Amazon Alexa) for health information [7].

Findings from the literature suggest that eHealth adoption is associated with health-protective outcomes among individuals. Previous studies purport that health IT adoption among individuals promotes healthy behaviors such as fruit and vegetable intake [8,9]. In addition, patients' use of online eHealth tools to communicate with providers is associated with increased access to health-related information, reduction of health care expenditures, improved emotional support, and enhanced clinical management and self-care [10-12]. Available findings also suggest that social media use among patients is associated with a positive impact on facilitating self-care [13-16]. With respect to prescription medication adherence, the use of eHealth tools to refill prescribed drugs is positively correlated with statin adherence among patients experiencing cardiovascular disease [17,18]. The US Department of Health and Human Services (HHS) recognizes the value that internet access and eHealth tools have on health outcomes, so they developed national goals in Healthy People 2020 to bolster access [19]. The HHS also created the Office of the National Coordinator, which has been advancing consumer eHealth via incentive programs, supporting developers to build eHealth tools, and strengthening trust and protecting privacy in health information technology (HIT) tools [20].

There is an emerging body of literature reporting the factors that influence eHealth tools adoption by consumers. Attributes related to geography, socioeconomic status, and gender are suggested to be associated with the utilization of eHealth tools. Geographic disparities in health IT uses were found between rural and urban residents. Specifically, rural residents were less likely to access online medical records [21]. Higher levels of education contribute to managing electronic personal health information (ePHI) online [22], OPPC [5], and use of health IT for self-regulation [8], which indicates that having a higher education is associated with better eHealth adoption and utilization. Gender has a mixed effect on eHealth tools adoption, depending on the scenario. Men with a risk of cardiovascular disease (CVD) have a higher likelihood of using wearable devices than people without CVD risk. However, there does not appear to be a difference among women [6]. Females had a significantly higher probability of using ePHI online management compared to males [22,23], but there is no significant difference in online patient-provider communication [5,23]. Income level was reported to be positively associated with eHealth sum score, a composite score representative of using technology to access health care [22,24].

Given that this study's focus population are Medicaid enrollees, a primary area of interest is the impact education and income have on individuals qualifying for this service. The inverse relationship between socioeconomic status and adverse health outcomes is well documented in public health literature. In the United States, increased household income is associated with longer life expectancy [25]. Though the causative nature of the relationship between cardiovascular disease and socioeconomic factors is yet to be fully understood, being classified in a lower socioeconomic stratum is linked to poor health outcomes such as hypertension, diabetes mellitus, and coronary heart disease [26]. Due to factors such as these and the unique challenges that financial hardship has on access to medical care [27], the United States government amended provisions in the Social Security Act, which established a health insurance program for impoverished citizens (ie, Medicaid). Findings with respect to internet access [28] parallel those of the underserved in the United States. Moreover, not to imply that being on Medicaid causes poor health outcomes, there is literature suggesting that Medicaid beneficiaries have poorer health outcomes that can lead to long-term cardiovascular complications compared to privately insured individuals [29].

Acknowledging the fact that one must have internet access to utilize eHealth tools and that national efforts should be made in the US to increase accessibility, there remain certain sectors of the population that are limited in this capacity. Nevertheless, given considerations of the potential health benefits that eHealth adoption may confer and current health disparities among Medicaid beneficiaries, a question we propose is “Are eHealth adoption rates among Medicaid beneficiaries comparable to those of people who are not on Medicaid?” To our best knowledge, no study has portrayed a comprehensive picture of eHealth tools adoption among adult Medicaid beneficiaries, an underserved population. Medicare and Medicaid EHR Incentive Programs have been promoting HIT adoption among providers and achieved a significant impact [20]. However, how Medicaid beneficiaries adopt eHealth tools was mostly unknown nationally. Due to the unique characteristics of Medicaid beneficiaries, the factors of eHealth adoption among the adult Medicaid population might also have different effects compared to those for the non-Medicaid community. Analyses from this study are focused on contrasting these characteristics between the two groups to provide research that may lead to health policy implications that positively impact America's underserved.

In this study, we use 3 years of the US Health Information National Trends Survey (HINTS) data to investigate the following research objectives (RO): RO1, comparison of eHealth tools adoption among Medicaid beneficiaries versus non-Medicaid insured individuals from 2017 to 2019; RO2, the association of Medicaid status with eHealth adoption; and RO3, compared to the non-Medicaid population, what factors are associated with increased eHealth adoption among the Medicaid population. To examine these questions, we assessed seven domains of eHealth tools utilization outcomes that covered a variety of HIT tools such as secure internet portals, EHRs,

patient-provider email messaging, mobile health apps, and wearable devices as described in the report [20].

Methods

Data and Sample

We extracted 3 waves of HINTS data that include the calendar years of 2017, 2018, and 2019. HINTS is a two-stage, cross-sectional, and nationally representative survey sampled from noninstitutionalized adults in the United States [30]. Since eHealth tools adoption is mainly dependent on internet access, HINTS 2017-2019 data operationalize internet access by participants who self-reported access to the internet in the following ways: regular dial-up telephone line, broadband, WiFi, and 3G or 4G cellular network. With this criterion, we excluded respondents who reported no internet access and obtained a sample size of 2513 respondents in 2017, 2704 respondents in 2018, and 4264 respondents in 2019.

Dependent Variables

We constructed seven primary outcomes that represented a variety of eHealth tools adoption and utilization behaviors. OPPC [5] is composed of three items: (1) used electronic means to communicate with a doctor, (2) whether tablet or smartphone helped discuss with health care providers, and (3) used online medical record to securely message health care providers. Health information management (HIM) [23] is composed of two items: (1) used electronic means to track health care charges and (2) used electronic means to look up medical test results. Mobile health for self-regulation (MHSR) [8] is composed of three items: (1) have apps related to health and wellness, (2) used electronic medical device to track health, and (3) used tablet or smartphone to help track progress on a health-related goal. Social media for health information (SMHI) [23] is composed of three items: (1) used internet to participate in an online forum or support group, (2) used internet to watch a health-related video on YouTube, and (3) used internet to share health information on social networking sites.

Sharing information [31] is composed of four items: (1) shared health information with a health professional via electronic device or smartphone, (2) electronically sent medical information to another health care provider, (3) electronically sent medical information to a family member or another person, and (4) electronically sent medical information to a service or app that helps manage and store health information. Buy or refill medicine (BRM) [32] is composed of two items: (1) used online medical record to request a refill of medications and (2) used electronic means to buy medicine or vitamins online. Decision making [31] is composed of two items: (1) used online medical record to help you make a decision about how to treat an illness or condition and (2) tablet or smartphone helped make a decision about how to treat an illness or condition. The definitions of each outcome are shown in Multimedia Appendix 1. The composite scores were constructed by summing their dichotomous items with responses “No” and “Yes” converted to 0 and 1, respectively, except that scores higher than 1 in SMHI were recoded to 2 due to sparsity of participants who share information via multiple channels.

Key Independent Variables and Covariates

Medicaid status is determined by participants' self-reported Medicaid insurance plan. Internet access diversity score was measured by whether participants had only one channel or multiple channels to access the internet. CVD risk was defined by having at least one of three conditions: hypertension, a heart condition, or diabetes. Depression was defined by whether respondents had been told they have depression or an anxiety disorder by a doctor or health professional. Demographic and socioeconomic variables included gender, age, race or ethnicity, education, residency, US Census region, and annual household income (Multimedia Appendix 2).

Statistical Analysis

We used descriptive statistics (frequencies and proportions) to demonstrate sample characteristics and eHealth tools adoption distribution patterns of HINTS data from 2017 to 2019. Pearson chi-square tests were performed to examine the association between Medicaid insurance status and other variables, including demographic, socioeconomic status, and eHealth outcomes. We considered a *P* value of less than or equal to .05 to be statistically significant. To answer RO2, we conducted ordinal logistic regressions for each of the seven eHealth outcomes using Medicaid status as the factor of primary interest. Base models included survey year as the only covariate, while adjusted models included the variables of age group, gender, education, Census region, residency, income, and internet access diversity. To answer RO3 and provide a comparison of detailed factor profiles in each population, we conducted ordinal logistic regression for each outcome within the Medicaid sample and non-Medicaid sample independently. In addition to demographic and socioeconomic factors, we included cardiovascular disease, depression, and internet access diversity as independent variables in the models. All statistical analyses except frequencies were conducted through jackknife weight procedures to represent national population-level estimates and obtain more precise confidence intervals. All analyses were performed in SAS 9.4 (SAS Institute).

Results

Table 1 presents the sample characteristics of both Medicaid and non-Medicaid population from 2017 to 2019. Compared to the people who were not enrolled in the Medicaid program, the Medicaid population is more likely to be young, belong to an underrepresented minority (Hispanic and non-Hispanic Black), and have low education and low family income. Generally, there is no significant difference in the distributions of residency between Medicaid and non-Medicaid populations. Interestingly, the proportions of people who have multiple ways to access the internet were not significantly different between Medicaid (68.6%) and non-Medicaid participants (67.6%), though the proportion of Medicaid beneficiaries who have no internet access at all was significantly higher than that of the non-Medicaid population (Multimedia Appendix 3).

Table 1. Demographic characteristics of sample from 3 iterations, 2017-2019, by Medicaid insurance status.

Variable	Medicaid (N=1087)	Non-Medicaid (N=8394)	*P* value[a]
Gender, n (%)			.01
Male	349 (42.1[b])	3545 (50.0)	—[c]
Female	716 (57.9)	4738 (50.0)	—
Education, n (%)			<.001
Less than high school	114 (11.3)	193 (4.0)	—
High school graduate	243 (29.7)	1039 (17.6)	—
Some college	445 (44.6)	2413 (39.2)	—
College graduate or higher	262 (14.4)	4621 (39.2)	—
Age group (years), n (%)			<.001
18-24	76 (20.3)	225 (9.4)	—
25-44	328 (36.1)	2108 (32.9)	—
45-64	475 (36.4)	3384 (41.9)	—
65+	181 (7.2)	2514 (15.8)	—
Race, n (%)			<.001
Hispanic	223 (22.7)	929 (13.5)	—
Non-Hispanic White	450 (51.3)	5439 (69.1)	—
Non-Hispanic Black	214 (15.8)	881 (8.9)	—
Non-Hispanic Other	66 (5.1)	274 (2.9)	—
Non-Hispanic Asian	48 (5.1)	350 (5.5)	—
Household income (US $), n (%)			<.001
Less than $20,000	528 (46.8)	564 (8.0)	—
$20,000 to <$35,000	200 (21.7)	801 (8.6)	—
$35,000 to <$50,000	129 (13.8)	991 (13.1)	—
$50,000 to <$75,000	80 (10.4)	1615 (20.9)	—
$75,000 or more	70 (7.2)	3666 (49.3)	—
Residency, n (%)			.82
Urban	970 (86.8)	7435 (87.2)	—
Rural	117 (13.2)	959 (12.8)	—
Census region, n (%)			<.001
Northeast	194 (17.8)	1252 (17.5)	—
Midwest	193 (21.2)	1503 (20.9)	—
South	365 (29.5)	3657 (38.6)	—
West	335 (31.5)	1982 (23.0)	—
Internet access diversity, n (%)			.67
One way to access internet	444 (31.2)	3289 (32.4)	—
More than one way to access	643 (68.8)	5105 (67.6)	—

[a]Chi-square tests were conducted to obtain *P* values adjusted for sampling weights.

[b]Percentages were weighted by jackknife weighting methods to represent US population-level estimates.

[c]Not available.

We investigated eHealth tools adoption rates among Medicaid and non-Medicaid populations. Table 2 shows the frequencies and weighted proportions of respondents by eHealth tools adoption scores. The higher scores suggest a better adoption and utilization of eHealth tools. Chi-square test results suggested no statistically significant disparity between Medicaid and

non-Medicaid respondents for sharing information, decision making, OPPC, or BRM, with exceptions for BRM in 2017 and OPPC in 2018. In the Medicaid population, the respondents were densely distributed in the score "0" of HIM with a range of 51.1% to 64.9%, which was 11.2% to 17.5% higher than that of the non-Medicaid population (P<.001). In 2018, 42.1% of Medicaid respondents had a 0 score for MHSR, which is 7.7% higher than non-Medicaid (P=.02). However, in 2019, the proportion of Medicaid respondents who reported a MHSR score of 0 dropped to 27.9%, only 0.8% higher than that of non-Medicaid. The proportions of Medicaid respondents were 7.8% and 9.4% less in scores 2 and 3 for MHSR compared to those of the non-Medicaid population (P<.001). Contrary to HIM and MHSR, Medicaid respondents had significantly higher SMHI adoption rates than non-Medicaid respondents in 2018 (8% higher, P=.01) and 2019 (10.1% higher, P=.01).

Table 2. The frequency and weighted prevalence of eHealth tools adoption among Medicaid and non-Medicaid populations from 2017 to 2019.

Variable	2017			2018			2019		
	M[a] (N=285)	NM[b] (N=2228)	*P* value[c]	M (N=299)	NM (N=2405)	*P* value	M (N=503)	NM (N=3761)	*P* value
SI[d] score, n (%)			.94			.81			.98
0	221 (81.2)	1738 (80.3)	—[e]	221 (80.5)	1816 (78.5)	—	363 (75.3)	2804 (75.9)	—
1	55 (16.2)	418 (17.3)	—	60 (15.7)	493 (18.0)	—	114 (20.5)	799 (19.9)	—
2	7 (2.6)	70 (2.3)	—	18 (3.8)	88 (3.5)	—	24 (4.3)	155 (4.3)	—
DM[f] score, n (%)			.69			.54			.32
0	162 (60.4)	1429 (62.6)	—	168 (60.9)	1400 (58.0)	—	250 (49.2)	2089 (56.0)	—
1	110 (37.4)	696 (33.9)	—	105 (32.4)	851 (36.9)	—	204 (43.1)	1328 (37.2)	—
2	10 (2.2)	85 (3.5)	—	22 (6.7)	137 (5.1)	—	38 (7.7)	303 (6.8)	—
HIM[g] score, n (%)			.01			<.001			.01
0	172 (61.2)	986 (44.3)	—	176 (64.9)	1029 (47.4)	—	261 (51.1)	1435 (39.9)	—
1	78 (30.8)	688 (31.0)	—	78 (25.9)	690 (26.6)	—	142 (29.4)	1182 (31.3)	—
2	33 (8.0)	549 (24.7)	—	41 (9.2)	663 (26.0)	—	100 (19.5)	1136 (28.8)	—
BRM[h] score, n (%)			<.001			.08			.46
0	218 (83.0)	1414 (67.0)	—	216 (74.7)	1492 (66.0)	—	307 (58.8)	2082 (56.6)	—
1	54 (14.0)	654 (27.0)	—	66 (20.9)	720 (27.6)	—	153 (33.4)	1235 (32.5)	—
2	10 (2.9)	156 (5.9)	—	17 (4.4)	189 (6.4)	—	41 (7.8)	438 (10.9)	—
MHSR[i] score, n (%)			.07			.02			<.001
0	117 (44.1)	789 (34.4)	—	121 (42.1)	802 (34.4)	—	142 (27.9)	1039 (27.1)	—
1	80 (26.3)	565 (24.4)	—	73 (23.6)	560 (20.1)	—	171 (37.4)	948 (21.0)	—
2	54 (18.7)	451 (22.1)	—	67 (22.7)	490 (22.6)	—	99 (16.6)	813 (24.4)	—
3	33 (10.9)	421 (19.1)	—	35 (11.6)	529 (22.9)	—	90 (18.1)	960 (27.5)	—
SMHI[j] score, n (%)			.14			.01			.01
0	137 (48.0)	1369 (57.7)	—	146 (48.0)	1385 (56.0)	—	224 (43.8)	2163 (53.9)	—
1	98 (40.3)	590 (29.2)	—	86 (29.4)	711 (32.2)	—	166 (33.0)	1150 (31.8)	—
2	28 (7.2)	201 (9.6)	—	45 (17.2)	214 (9.3)	—	79 (16.5)	320 (10.4)	—
3	20 (4.4)	65 (3.5)	—	18 (5.4)	73 (2.5)	—	34 (6.7)	118 (3.9)	—
OPPC[k] score, n (%)			.21			<.001			.92
0	155 (53.8)	966 (44.9)	—	159 (62.8)	995 (44.9)	—	217 (38.0)	1349 (39.1)	—
1	71 (27.6)	634 (28.7)	—	76 (19.1)	669 (25.7)	—	139 (28.7)	993 (26.0)	—
2	37 (10.8)	410 (18.2)	—	33 (9.6)	453 (18.4)	—	89 (20.3)	854 (21.0)	—
3	21 (7.8)	215 (8.2)	—	27 (8.6)	276 (11.0)	—	57 (12.9)	562 (13.9)	—

[a]M: Medicaid population.

[b]NM: non-Medicaid population.

[c]Chi-square tests were conducted to obtain *P* values adjusted for sampling weights.

[d]SI: sharing information.

[e]Not available.

[f]DM: decision making.

[g]HIM: health information management.

[h]BRM: buy or refill medicine.

[i]MHSR: mobile health for self-regulation.

[j]SMHI: social media for health information.

[k]OPPC: online patient-provider communication.

Multimedia Appendix 4 shows that the effects of Medicaid status on BRM and OPPC were attenuated after adjusting multiple covariates. Consistent with Table 2, Medicaid beneficiaries were significantly less likely to use eHealth tools for HIM (OR 0.64, 95% CI 0.50-0.82) and MHSR (OR 0.63, 95% CI 0.53-0.76), but more likely for SMHI (OR 1.35, 95% CI 1.12-1.63). To address RO3, we further assessed the factor effects such as demographic, socioeconomic, health, and internet access on these three eHealth categories in the Medicaid and non-Medicaid populations independently. Table 3 and 4 shows that in the non-Medicaid population, HIM and MHSR were more likely to be adopted by individuals who were female, were well educated, had high-income households, and had more than one source of internet access. Having cardiovascular disease or depression was also significantly associated with HIM and MHSR. Among the non-Medicaid respondents, the SMHI was significantly associated with factors such as gender, race, age, depression, and internet access diversity (Table 5). However, there were fewer factors significantly associated with eHealth tools adoption among the Medicaid respondents. Education (P=.03), CVD risk (OR 2.16, 95% CI 1.18-3.93; P=.01), depression (OR 1.67, 95% CI 1.07-2.6; P=.02), and internet access diversity (OR 2.87, 95% CI 1.55-5.30; P=.001) were factors significantly associated with HIM (Table 3). Education (P=.02), CVD risk (OR 2.55, 95% CI 1.52-4.3; P<.001) and internet access diversity (OR 2.72, 95% CI 1.69-4.38; P<.001) were factors significantly associated with MHSR (Table 4). Internet access (OR 1.98, 95% CI 1.26-3.11; P=.004) was the only significant factor associated with SMHI (Table 5).

Table 3. The odds ratio of predictors for health information management via eHealth tools among Medicaid and non-Medicaid participants.

Variable	Medicaid		Non-Medicaid	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Year				
2017	Ref	.01	Ref	.01
2018	1.23 (0.68-2.22)	—[b]	0.95 (0.79-1.15)	—
2019	2.22 (1.29-3.80)	—	1.24 (1.03-1.49)	—
Gender				
Male	Ref	.83	Ref	<.001
Female	0.95 (0.57-1.58)	—	1.36 (1.15-1.60)	—
Race				
Non-Hispanic White	Ref	.80	Ref	.02
Hispanic	1.43 (0.72-2.84)	—	0.92 (0.69-1.22)	—
Non-Hispanic Black	1.31 (0.72-2.37)	—	0.69 (0.51-0.95)	—
Non-Hispanic Other	0.83 (0.32-2.16)	—	1.10 (0.65-1.86)	—
Non-Hispanic Asian	1.31 (0.37-4.65)	—	1.59 (1.04-2.44)	—
Education				
Less than high school	Ref	.03	Ref	<.001
High school graduate	0.84 (0.34-2.09)	—	1.71 (0.95-3.07)	—
Some college	1.68 (0.68-4.14)	—	2.85 (1.56-5.23)	—
College graduate or higher	2.33 (0.91-5.92)	—	4.37 (2.48-7.70)	—
Age group (years)				
18-24	Ref	.72	Ref	.06
25-44	1.58 (0.66-3.80)	—	1.54 (1.01-2.35)	—
45-64	1.46 (0.55-3.85)	—	1.37 (0.91-2.06)	—
65+	1.63 (0.55-4.86)	—	1.18 (0.76-1.82)	—
Census region				
Northeast	Ref	.55	Ref	.03
Midwest	1.36 (0.70-2.66)	—	1.22 (0.95-1.56)	—
South	1.50 (0.69-3.27)	—	1.21 (0.93-1.58)	—
West	0.95 (0.46-1.99)	—	1.46 (1.14-1.87)	—
Residency				
Urban	Ref	.90	Ref	.01
Rural	0.96 (0.49-1.87)	—	0.71 (0.55-0.90)	—
Household income (US $)				
Less than $20,000	Ref	.23	Ref	<.001
$20,000 to <$35,000	0.85 (0.47-1.55)	—	1.09 (0.69-1.73)	—
$35,000 to <$50,000	1.76 (0.95-3.25)	—	1.48 (0.89-2.44)	—
$50,000 to <$75,000	1.40 (0.68-2.91)	—	1.56 (0.96-2.54)	—
$75,000 or more	1.79 (0.70-4.56)	—	1.94 (1.20-3.13)	—
Cardiovascular disease				
No	Ref	.01	Ref	.01
Yes	2.16 (1.18-3.93)	—	1.29 (1.07-1.56)	—

Variable	Medicaid		Non-Medicaid	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Depression				
No	Ref	.02	Ref	<.001
Yes	1.67 (1.07-2.60)	—	1.41 (1.18-1.68)	—
Internet access diversity				
One way to access internet	Ref	.001	Ref	<.001
More than one way to access	2.87 (1.55-5.30)	—	1.45 (1.18-1.78)	—

[a]OR: odds ratio.

[b]Not available.

Table 4. The odds ratio of predictors for mobile health for self-regulation among Medicaid and non-Medicaid participants.

Variable	Medicaid		Non-Medicaid	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Year				
2017	Ref	.10	Ref	<.001
2018	1.28 (0.73-2.24)	—[b]	1.18 (0.96-1.44)	—
2019	1.73 (1.04-2.88)	—	1.49 (1.24-1.80)	—
Gender				
Male	Ref	.21	Ref	<.001
Female	1.37 (0.83-2.26)	—	1.56 (1.35-1.80)	—
Race				
Non-Hispanic White	Ref	.57	Ref	.07
Hispanic	1.47 (0.85-2.54)	—	1.14 (0.92-1.42)	—
Non-Hispanic Black	1.35 (0.57-3.19)	—	1.16 (0.88-1.55)	—
Non-Hispanic Other	1.55 (0.45-5.27)	—	1.37 (0.97-1.92)	—
Non-Hispanic Asian	0.78 (0.33-1.81)	—	1.32 (0.86-2.01)	—
Education				
Less than high school	Ref	.02	Ref	<.001
High school graduate	1.14 (0.44-2.99)	—	0.97 (0.61-1.54)	—
Some college	0.78 (0.31-1.97)	—	1.74 (1.18-2.57)	—
College graduate or higher	1.82 (0.65-5.11)	—	2.04 (1.38-3.02)	—
Age group (years)				
18-24	Ref	.37	Ref	<.001
25-44	0.59 (0.23-1.50)	—	0.94 (0.65-1.36)	—
45-64	0.47 (0.18-1.26)	—	0.63 (0.43-0.93)	—
65+	0.45 (0.18-1.15)	—	0.41 (0.28-0.60)	—
Census region				
Northeast	Ref	.99	Ref	.03
Midwest	1.02 (0.50-2.09)	—	1.23 (0.97-1.55)	—
South	0.94 (0.46-1.89)	—	1.36 (1.11-1.67)	—
West	1.03 (0.47-2.30)	—	1.22 (1.01-1.47)	—
Residency				
Urban	Ref	.97	Ref	.51
Rural	1.01 (0.51-2.03)	—	0.93 (0.75-1.16)	—
Household income (US $)				
Less than $20,000	Ref	.44	Ref	<.001
$20,000 to <$35,000	1.27 (0.75-2.17)	—	1.24 (0.79-1.94)	—
$35,000 to <$50,000	1.28 (0.56-2.89)	—	1.66 (1.08-2.54)	—
$50,000 to <$75,000	1.95 (0.75-5.08)	—	1.73 (1.17-2.55)	—
$75,000 or more	2.14 (0.80-5.69)	—	2.81 (1.88-4.22)	—
Cardiovascular disease				
No	Ref	<.001	Ref	<.001
Yes	2.55 (1.52-4.30)	—	1.44 (1.20-1.72)	—

Variable	Medicaid		Non-Medicaid	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Depression				
No	Ref	.18	Ref	.63
Yes	1.34 (0.87-2.05)	—	1.05 (0.85-1.30)	—
Internet access diversity				
One way to access internet	Ref	<.001	Ref	<.001
More than one way to access	2.72 (1.69-4.38)	—	1.47 (1.29-1.68)	—

[a]OR: odds ratio.

[b]Not available.

Table 5. The odds ratio of predictors for social media for health information among Medicaid and non-Medicaid participants.

Variable	Medicaid		Non-Medicaid	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Year				
2017	Ref	.07	Ref	.13
2018	1.79 (1.04-3.10)	—[b]	1.08 (0.87-1.35)	—
2019	1.95 (1.05-3.63)	—	1.19 (1.00-1.43)	—
Gender				
Male	Ref	.18	Ref	<.001
Female	1.36 (0.86-2.14)	—	1.50 (1.30-1.72)	—
Race				
Non-Hispanic White	Ref	.06	Ref	<.001
Hispanic	1.37 (0.72-2.63)	—	1.44 (1.13-1.83)	—
Non-Hispanic Black	1.54 (0.82-2.89)	—	1.30 (0.96-1.75)	—
Non-Hispanic Other	3.57 (1.28-9.97)	—	1.37 (0.90-2.08)	—
Non-Hispanic Asian	1.86 (0.82-4.24)	—	1.98 (1.39-2.80)	—
Education				
Less than high school	Ref	.21	Ref	.09
High school graduate	0.70 (0.32-1.51)	—	1.01 (0.54-1.91)	—
Some college	0.95 (0.45-2.04)	—	1.39 (0.83-2.33)	—
College graduate or higher	1.27 (0.59-2.72)	—	1.39 (0.83-2.35)	—
Age group (years)				
18-24	Ref	.06	Ref	<.001
25-44	1.29 (0.61-2.73)	—	0.87 (0.58-1.32)	—
45-64	1.45 (0.68-3.06)	—	0.57 (0.39-0.84)	—
65+	0.59 (0.23-1.50)	—	0.30 (0.21-0.42)	—
Census region				
Northeast	Ref	.16	Ref	.09
Midwest	1.96 (1.03-3.72)	—	0.98 (0.76-1.28)	—
South	1.73 (0.92-3.26)	—	1.26 (0.99-1.61)	—
West	1.67 (0.88-3.15)	—	1.24 (0.96-1.59)	—
Residency				
Urban	Ref	.44	Ref	.16
Rural	0.71 (0.29-1.75)	—	0.81 (0.59-1.09)	—
Household income (US $)				
Less than $20,000	Ref	.97	Ref	.54
$20,000 to <$35,000	0.96 (0.50-1.87)	—	1.42 (0.94-2.16)	—
$35,000 to <$50,000	1.15 (0.67-1.95)	—	1.31 (0.87-1.98)	—
$50,000 to <$75,000	0.94 (0.42-2.11)	—	1.17 (0.84-1.64)	—
$75,000 or more	0.99 (0.45-2.16)	—	1.22 (0.87-1.72)	—
Cardiovascular disease				
No	Ref	.85	Ref	.47
Yes	1.04 (0.68-1.58)	—	1.07 (0.89-1.29)	—

Variable	Medicaid		Non-Medicaid	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value
Depression				
No	Ref	.51	Ref	<.001
Yes	1.18 (0.71-1.97)	—	1.54 (1.28-1.86)	—
Internet access diversity				
One way to access internet	Ref	.004	Ref	.01
More than one way to access	1.98 (1.26-3.11)	—	1.33 (1.09-1.61)	—

[a]OR: odds ratio.

[b]Not available.

Discussion

According to Ambrosi and colleagues [33], the digital divide can be operationalized in multiple forms. The most common way of defining the digital divide relates to the inequalities of ICT access, for example, complete lack of access to the internet or a smart device. Findings from this study suggest that the Medicaid population is at higher odds of having no access to the internet when compared to the non-Medicaid population. Though worthy of attention and immediate addressing, the elimination or significant reduction of this type of disparity will be an undertaking that will likely not happen in the near future. The second form of the digital divide is caused by different use patterns among individuals who already have access to ICT. An example is people who only use the basic functions (talk and texting) within a smartphone that has many more available features. The key objectives of our study focused on the latter scenario and aimed at assessing the profiles and patterns of digital inequality between the Medicaid and the non-Medicaid populations with some form of internet access on a national level in the United States. We found that the Medicaid population had lower adoption on HIM and MHSR than the non-Medicaid population, which is consistent with a previous study about the effect of socioeconomic status [34]. According to Venkatesh and colleagues' unified theory of acceptance and use of technology (UTAUT), effort expectancy (degree of ease of use of technology) and social influence are important direct factors that are positively associated with the intention and behavior of consumers' technology adoption [35]. Underserved populations like that of Medicaid beneficiaries may perceive eHealth tools to be a more challenging resource to adopt due to lower education as well as eHealth literacy, the ability to access, process, understand, and use the features of the technology [36]. Social influence may also play a critical role in limiting Medicaid participants' use of smart devices to full capacity, though the effect of social influence has not been widely tested among the Medicaid population. Interestingly, we found that the Medicaid population was more likely to use social media to engage in health-related activities. The barriers for underserved communities to regularly access medical resources may push Medicaid participants to use social media as a cost-effective alternative source of medical information. This increased adoption of social media use among the Medicaid population could be a potential double-edged sword. On the one hand, the low quality of medical information on social media may mislead patients to avoid seeking treatment from regular health professionals. Conversely, well-designed social media interventions that specifically target underserved populations could be developed to promote eHealth technology adoption and eHealth literacy through social media. The challenge of this approach is that current social media artificial intelligence and machine learning algorithms may not have the ability to differentiate social media posts that are beneficial and detrimental to users' understanding of health. Therefore, increased exposure to social media interventions may also increase exposure to misleading social media posts.

Several studies have examined the predictors of eHealth tasks or tools adoption among the general population [8,34,37]. Being female, being younger, and having higher income was consistently associated with higher eHealth adoption in general populations, which is consistent with our results based on the non-Medicaid population. However, these factors were not significantly associated with any of the three major eHealth categories: HIM, MHSR, and SMHI. The attenuation of these factor effects in underserved populations was also reported in several studies [38,39], but none has compared the factor effects of underserved populations to their counterparts, nor have they explained the reason. It is possible that the use of eHealth tools among underserved populations like Medicaid beneficiaries was determined by other key factors (eg, effort expectancy and social influence), of which the levels do not vary much across subgroups of gender, age, or income level. Unlike income, attained education was significantly associated with both HIM and MHSR, and the effects were not monotonically increasing with higher education level. Therefore, education may not be the factor directly influencing eHealth tools adoption among this population, or the effect of education was confounded by hidden factors. Medicaid beneficiaries with CVD risk were significantly more likely to engage in HIT and use an eHealth app for health benefits. On the other hand, Medicaid participants with depression demonstrated significantly higher odds only in HIM. The adoption of MHSR involves more self-discipline and motivation, which might be a hurdle for depressive Medicaid respondents to use their smart devices for health-related purposes. This also leaves room for developing interventions or tools to facilitate depressive patients to monitor their mental health and improve their adoption of eHealth tools to monitor their physical health as well. Finally, findings suggest that

enhancing internet access diversity was significantly associated with improving HIM, MHSR, and SMHI among both Medicaid and non-Medicaid groups. The odds ratios of shifting internet access from a single source to multiple sources were almost doubled among the Medicaid population compared to their counterparts. Internet access diversity, aligned with infrastructure and supports, may determine the facilitating conditions which directly influence the technology use behaviors based on UTAUT [35]. Thus, providing underserved populations with more sources to access the internet or smart devices (eg, cheap 4G data plan, free public WiFi hot spot, affordable Android smartphone) could be a cost-effective means to improve their eHealth adoption and use. In fact, the Lifeline program, which has been providing discounts on phone service for low-income consumers starting in 1985, has adopted broadband (3G) as a support service since 2016 [40]. As such, increasing awareness of the Federal Communications Commission's Lifeline program is an immediate short-term goal to increase access to smartphones and broadband internet for families eligible for Medicaid.

We believe our study has several strengths. To our best knowledge, this study is the first to investigate eHealth technology and tools adoption among underserved populations, especially the Medicaid enrollees, at the national level in the United States. Therefore, sample selection bias was greatly reduced compared to studies that recruited a small sample size of local participants [38,39]. Additionally, we assessed eHealth technology and tools adoption using seven composite scores constructed from 19 items, which provides a comprehensive picture and reduces measurement errors from using single items. Finally, providing increased attention on this topic will hopefully open a new area of research that may have profound effects on the nation's health and economy. Given the above important contributions to the literature, we acknowledge the limitations of the study. Like other studies using HINTS data sets, our study is based on cross-sectional surveys. Therefore, it could not account for all confounding variables or evaluate causal relationships. The questionnaires designed by HINTS were not suitable to investigate the psychometric constructs that impact eHealth technology adoption directly. In addition, the HINTS data are based on self-reported data from respondents. We are unable to confirm if their answers to the eHealth-related questions were without errors. We found a small portion of respondents reported never accessing the internet, but they also reported a positive response to eHealth technology utilization, though we have excluded those respondents to eliminate spuriousness.

To effectively reduce the disparities in eHealth tools adoption between underserved populations and their counterparts, certain factors must be taken into account when developing interventions or infrastructures for Medicaid beneficiaries. Our study offers initial insights into the factors among the nation’s underserved population. Nevertheless, the fundamental drivers of eHealth use among the Medicaid population may not be fully revealed yet. In addition to confirming this study's findings, future studies should take our investigation to a more granular level and examine the intention and behaviors of eHealth tools utilization among underserved populations under the framework of existing models such as UTAUT. Additionally, the effects of interventions aimed at improving eHealth utilization (eg, improving internet access diversity) could be studied more closely to further validate our findings.

Conflicts of Interest

None declared.

Multimedia Appendix 1
eHealth technologies and tools adoption outcomes and definitions.
[DOCX File , 14 KB - jmir_v23i2e25809_app1.docx]

Multimedia Appendix 2
Definitions of independent variables.
[DOCX File , 14 KB - jmir_v23i2e25809_app2.docx]

Multimedia Appendix 3
The number and weighted percentage of respondents with no Internet access, one way to access Internet and multiple ways to access Internet.
[DOCX File , 14 KB - jmir_v23i2e25809_app3.docx]

Multimedia Appendix 4
Odds ratio (OR) and 95% confidence interval (CI) of Medicaid insurance status by the base model and adjusted model.
[DOCX File , 14 KB - jmir_v23i2e25809_app4.docx]

References

1. Serrano K, Thai C, Greenberg A, Blake K, Moser R, Hesse B. Progress on Broadband Access to the Internet and Use of Mobile Devices in the United States. Public Health Rep 2017;132(1):27-31 [FREE Full text] [doi: 10.1177/0033354916679365] [Medline: 28005473]

2. Bauerly B, McCord R, Hulkower R, Pepin D. Broadband Access as a Public Health Issue: The Role of Law in Expanding Broadband Access and Connecting Underserved Communities for Better Health Outcomes. J Law Med Ethics 2019 Jun;47(2_suppl):39-42 [FREE Full text] [doi: 10.1177/1073110519857314] [Medline: 31298126]
3. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb 24;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
4. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a Conceptual Model for eHealth: Qualitative Study with Key Informants. J Med Internet Res 2017 Oct 24;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]
5. Jiang S, Hong Y, Liu P. Trends of online patient-provider communication among cancer survivors from 2008 to 2017: a digital divide perspective. J Cancer Surviv 2019 Apr;13(2):197-204. [doi: 10.1007/s11764-019-00742-4] [Medline: 30756225]
6. Shan R, Ding J, Plante TB, Martin SS. Mobile Health Access and Use Among Individuals With or At Risk for Cardiovascular Disease: 2018 Health Information National Trends Survey (HINTS). J Am Heart Assoc 2019 Dec 17;8(24):e014390 [FREE Full text] [doi: 10.1161/JAHA.119.014390] [Medline: 31818220]
7. Alagha E, Helbing R. Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: an exploratory comparison of Alexa, Google Assistant and Siri. BMJ Health Care Inform 2019 Nov;26(1):a [FREE Full text] [doi: 10.1136/bmjhci-2019-100075] [Medline: 31767629]
8. Li L, Peng W. Does Health Information Technology Promote Healthy Behaviors? The Mediating Role of Self-Regulation. Health Commun 2020 Dec;35(14):1772-1781. [doi: 10.1080/10410236.2019.1663468] [Medline: 31496292]
9. Chae J. How we use the Internet matters for health: The relationship between various online health-related activities and preventive dietary behaviors. Health Informatics J 2019 Sep;25(3):973-983 [FREE Full text] [doi: 10.1177/1460458217735675] [Medline: 29047292]
10. Santana S, Lausen B, Bujnowska-Fedak M, Chronaki C, Kummervold PE, Rasmussen J, et al. Online communication between doctors and patients in Europe: status and perspectives. J Med Internet Res 2010 Jun 15;12(2):e20 [FREE Full text] [doi: 10.2196/jmir.1281] [Medline: 20551011]
11. Tarver W, Menser T, Hesse B, Johnson T, Beckjord E, Ford E, et al. Growth Dynamics of Patient-Provider Internet Communication: Trend Analysis Using the Health Information National Trends Survey (2003 to 2013). J Med Internet Res 2018 Mar 29;20(3):e109 [FREE Full text] [doi: 10.2196/jmir.7851] [Medline: 29599107]
12. McGrory A. Communicating with head and neck cancer patients. ORL Head Neck Nurs 2011;29(3):7-11. [Medline: 21902107]
13. Hamm M, Chisholm A, Shulhan J, Milne A, Scott S, Given L, et al. Social media use among patients and caregivers: a scoping review. BMJ Open 2013 May 09;3(5):e002819 [FREE Full text] [doi: 10.1136/bmjopen-2013-002819] [Medline: 23667163]
14. Fat MJL, Doja A, Barrowman N, Sell E. YouTube videos as a teaching tool and patient resource for infantile spasms. J Child Neurol 2011 Jul;26(7):804-809. [doi: 10.1177/0883073811402345] [Medline: 21551373]
15. Merkel RM, Wright T. Parental self-efficacy and online support among parents of children diagnosed with type 1 diabetes mellitus. Pediatr Nurs 2012;38(6):303-8; quiz 309. [Medline: 23362628]
16. Hamm MP, Shulhan J, Williams G, Milne A, Scott SD, Hartling L. A systematic review of the use and effectiveness of social media in child health. BMC Pediatr 2014 Jun 02;14:138 [FREE Full text] [doi: 10.1186/1471-2431-14-138] [Medline: 24886048]
17. Sarkar U, Lyles CR, Parker MM, Allen J, Nguyen R, Moffet HH, et al. Use of the refill function through an online patient portal is associated with improved adherence to statins in an integrated health system. Med Care 2014 Mar;52(3):194-201 [FREE Full text] [doi: 10.1097/MLR.0000000000000069] [Medline: 24374412]
18. Lyles C, Sarkar U, Schillinger D, Ralston J, Allen J, Nguyen R, et al. Refilling medications through an online patient portal: consistent improvements in adherence across racial/ethnic groups. J Am Med Inform Assoc 2016 Apr;23(e1):e28-e33 [FREE Full text] [doi: 10.1093/jamia/ocv126] [Medline: 26335983]
19. HC/HIT-6.2 Data Details | Healthy People 2020. 2020. URL: https://www.healthypeople.gov/node/4540/data_details [accessed 2021-02-08]
20. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
21. Krakow M, Hesse B, Oh A, Patel V, Vanderpool R, Jacobsen P. Addressing Rural Geographic Disparities Through Health IT: Initial Findings From the Health Information National Trends Survey. Med Care 2019 Jun;57 Suppl 6 Suppl 2:S127-S132. [doi: 10.1097/MLR.0000000000001028] [Medline: 31095051]
22. Greenberg-Worisek A, Kurani S, Finney Rutten L, Blake K, Moser R, Hesse B. Assessing electronic personal health information use: An update on progress toward healthy people 2020 objectives. Health Policy and Technology 2019 Sep;8(3):211-220. [doi: 10.1016/j.hlpt.2019.05.010]
23. Jackson D, Chou W, Coa K, Oh A, Hesse B. Implications of social media use on health information technology engagement: Data from HINTS 4 cycle 3. Transl Behav Med 2016 Dec;6(4):566-576 [FREE Full text] [doi: 10.1007/s13142-016-0437-1] [Medline: 27628436]

24. Sherman L, Patterson M, Tomar A, Wigfall L. Use of Digital Health Information for Health Information Seeking Among Men Living With Chronic Disease: Data From the Health Information National Trends Survey. Am J Mens Health 2020;14(1):1557988320901377 [FREE Full text] [doi: 10.1177/1557988320901377] [Medline: 31973642]
25. Chetty R, Stepner M, Abraham S, Lin S, Scuderi B, Turner N, et al. The Association Between Income and Life Expectancy in the United States, 2001-2014. JAMA 2016 Apr 26;315(16):1750-1766 [FREE Full text] [doi: 10.1001/jama.2016.4226] [Medline: 27063997]
26. Schultz W, Kelli H, Lisko J, Varghese T, Shen J, Sandesara P, et al. Socioeconomic Status and Cardiovascular Outcomes: Challenges and Interventions. Circulation 2018 May 15;137(20):2166-2178 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.117.029652] [Medline: 29760227]
27. Lazar M, Davenport L. Barriers to Health Care Access for Low Income Families: A Review of Literature. J Community Health Nurs 2018;35(1):28-37. [doi: 10.1080/07370016.2018.1404832] [Medline: 29323941]
28. Karpman M. Many Adults Targeted by Medicaid Work Requirements Face Barriers to Sustained Employment. Washington, DC: Urban Institute; 2019. URL: https://www.urban.org/sites/default/files/publication/100296/hrms_medicaid_work_requirements.pdf [accessed 2021-02-08]
29. Mylona E, Benitez G, Shehadeh F, Fleury E, Mylonakis S, Kalligeros M, et al. The association of obesity with health insurance coverage and demographic characteristics: a statewide cross-sectional study. Medicine (Baltimore) 2020 Jul 02;99(27):e21016 [FREE Full text] [doi: 10.1097/MD.0000000000021016] [Medline: 32629722]
30. Survey Instruments. Health Information National Trends Survey (HINTS). URL: https://hints.cancer.gov/data/survey-instruments.aspx [accessed 2020-07-30]
31. Langford A, Orellana K, Kalinowski J, Aird C, Buderer N. Use of Tablets and Smartphones to Support Medical Decision Making in US Adults: Cross-Sectional Study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19531 [FREE Full text] [doi: 10.2196/19531] [Medline: 32784181]
32. McInnes D, Shimada S, Midboe A, Nazi K, Zhao S, Wu J, et al. Patient Use of Electronic Prescription Refill and Secure Messaging and Its Association With Undetectable HIV Viral Load: A Retrospective Cohort Study. J Med Internet Res 2017 Feb 15;19(2):e34 [FREE Full text] [doi: 10.2196/jmir.6932] [Medline: 28202428]
33. Ambrosi A, Peugeot V, Pimienta D. Word Matters: multicultural perspectives on information societies. C & F (edn) 2005.
34. Kontos E, Blake K, Chou W, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
35. Venkatesh V, Thong JYL, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]
36. Kumar D, Hemmige V, Kallen M, Giordano T, Arya M. Mobile Phones May Not Bridge the Digital Divide: A Look at Mobile Phone Literacy in an Underserved Patient Population. Cureus 2019 Feb 20;11(2):e4104 [FREE Full text] [doi: 10.7759/cureus.4104] [Medline: 31057998]
37. Hong Y, Jiang S, Liu P. Use of Patient Portals of Electronic Health Records Remains Low From 2014 to 2018: Results From a National Survey and Policy Implications. Am J Health Promot 2020 Jul;34(6):677-680. [doi: 10.1177/0890117119900591] [Medline: 32030989]
38. Schnall R, Cho H, Webel A. Predictors of willingness to use a smartphone for research in underserved persons living with HIV. Int J Med Inform 2017 Mar;99:53-59 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.01.002] [Medline: 28118922]
39. Nambisan P. Factors that impact Patient Web Portal Readiness (PWPR) among the underserved. Int J Med Inform 2017 Jun;102:62-70. [doi: 10.1016/j.ijmedinf.2017.03.004] [Medline: 28495349]
40. Federal Communications Commission. Lifeline Program for Low-Income Consumers Internet. 2012. URL: https://www.fcc.gov/general/lifeline-program-low-income-consumers [accessed 2021-02-08]

Abbreviations

CVD: cardiovascular disease
EHR: electronic health record
ePHI: electronic personal health information
HHS: US Department of Health and Human Services
HIM: health information management
HINTS: Health Information National Trends Survey
HIT: health information technology
ICT: information and communication technologies
MHSR: mobile health for self-regulation
OPPC: online patient-provider communication
RO: research objective
SMHI: social media for health information
UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 17.11.20; peer-reviewed by A Langford; comments to author 10.12.20; revised version received 22.12.20; accepted 20.01.21; published 18.02.21.

Please cite as:
Yang X, Yang N, Lewis D, Parton J, Hudnall M
Patterns and Influencing Factors of eHealth Tools Adoption Among Medicaid and Non-Medicaid Populations From the Health Information National Trends Survey (HINTS) 2017-2019: Questionnaire Study
J Med Internet Res 2021;23(2):e25809
URL: http://www.jmir.org/2021/2/e25809/
doi:10.2196/25809
PMID:33599619

Original Paper

A Social Media Campaign (#datasaveslives) to Promote the Benefits of Using Health Data for Research Purposes: Mixed Methods Analysis

Lamiece Hassan[1], PhD; Goran Nenadic[2*], PhD; Mary Patricia Tully[3*], BSc, MSc, PhD, FRPharmS, FFRPS

[1]Division of Informatics, Imaging and Data Sciences, The University of Manchester, Manchester, United Kingdom

[2]Department of Computer Science, The University of Manchester, Manchester, United Kingdom

[3]Division of Pharmacy and Optometry, The University of Manchester, Manchester, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Lamiece Hassan, PhD
Division of Informatics, Imaging and Data Sciences
The University of Manchester
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 01612751160
Email: lamiece.hassan@manchester.ac.uk

Abstract

Background: Social media provides the potential to engage a wide audience about scientific research, including the public. However, little empirical research exists to guide health scientists regarding what works and how to optimize impact. We examined the social media campaign #datasaveslives established in 2014 to highlight positive examples of the use and reuse of health data in research.

Objective: This study aims to examine how the #datasaveslives hashtag was used on social media, how often, and by whom; thus, we aim to provide insights into the impact of a major social media campaign in the UK health informatics research community and further afield.

Methods: We analyzed all publicly available posts (tweets) that included the hashtag #datasaveslives (N=13,895) on the microblogging platform Twitter between September 1, 2016, and August 31, 2017. Using a combination of qualitative and quantitative analyses, we determined the frequency and purpose of tweets. Social network analysis was used to analyze and visualize tweet sharing (*retweet*) networks among hashtag users.

Results: Overall, we found 4175 original posts and 9720 retweets featuring #datasaveslives by 3649 unique Twitter users. In total, 66.01% (2756/4175) of the original posts were retweeted at least once. Higher frequencies of tweets were observed during the weeks of prominent policy publications, popular conferences, and public engagement events. Cluster analysis based on retweet relationships revealed an interconnected series of groups of #datasaveslives users in academia, health services and policy, and charities and patient networks. Thematic analysis of tweets showed that #datasaveslives was used for a broader range of purposes than indexing information, including event reporting, encouraging participation and action, and showing personal support for data sharing.

Conclusions: This study shows that a hashtag-based social media campaign was effective in encouraging a wide audience of stakeholders to disseminate positive examples of health research. Furthermore, the findings suggest that the campaign supported community building and bridging practices within and between the interdisciplinary sectors related to the field of health data science and encouraged individuals to demonstrate personal support for sharing health data.

(J Med Internet Res 2021;23(2):e16348) doi:10.2196/16348

KEYWORDS

social media; public engagement; social network analysis; medical research

Introduction

Social Media Use by Academics

Social media platforms such as Twitter, LinkedIn, and Facebook have changed the way scientists interact with others, both socially and professionally. Although the specifics may vary between individuals, platforms, and scientific disciplines [1], common scholarly purposes for using social media among academics include discovering peers and enhancing collaboration, sharing links or citations to their own or others work, communicating the proceedings of conferences and meetings, raising their own profiles, engaging in discussions and keeping up to date with scholarly work, answering questions and solving problems, and discovering job opportunities [2-13].

There is also a growing interest in using social media to engage a wider audience about scientific research, including the public [6,14,15]. A recent scoping review of health scientists' strategies by Fontaine et al [16] identified 9 types of science communication strategies used by health scientists, directed at areas such as content, engagement, intention, presentation, and statistics. However, the same review concluded that empirical studies in this field were lacking, representing a missed opportunity to understand how to optimize science communication strategies.

A Social Media Campaign for Health Informatics Research: #datasaveslives

The social media campaign #datasaveslives [17] was established in 2014 by the Northern England branch of the Farr Institute for Health Informatics Research, a publicly funded, UK-wide research collaboration involving academic institutions and health partners. The campaign started with a simple goal: to promote the positive use of data in health research on social media. A select group of academic organizations belonging to, or affiliated with, the Farr Institute subsequently formally adopted #datasaveslives as part of their communications and stakeholder engagement strategies [18]. These supporters then encouraged a wider audience of people who supported health data research to use the hashtag #datasaveslives on social media sites (primarily Twitter) to index and share examples that demonstrate how health data from patient records and other sources could be used to create public health benefits. The second objective is to spark interest and dialog about using health data for research purposes among wider audiences, including patients, members of the public, health care professionals, and policy makers.

About Twitter

Twitter is a popular microblogging social media platform founded in 2006 [19]. It allows users to post short messages (previously 140 characters, more recently extended to 280) known as *tweets*, which may also include URL links, multimedia content (eg, images or videos) and/or references to other users (signified using the @ symbol, plus a username). Hashtags may also be used by assigning the # character to a term of their choice; this is a useful way of indexing and searching for tweets on a similar topic. Users can view and engage with tweets in a number of ways, including liking, replying to, and sharing (*retweeting*) others' posts. They can also follow others to *subscribe* to see their tweets. Tweets are public by default, although users can change their settings at any time to restrict their visibility to their Twitter followers. Users can also choose to write a short description about themselves (known as a *bio*) and add their location.

Study Aim

The aim of this study is to examine how the #datasaveslives hashtag has been used on Twitter in the context of the use of data in health research and by whom. The analysis will determine how often the hashtag has been used and shared and examine the content posted alongside the hashtag to determine the range of purposes for its use. This will provide insights into the strategic use of social media campaigns by academics and explore their potential for encouraging wider dialog within and between scientific communities and broader audiences.

Specifically, the following objectives (and research questions in brackets) were defined:

1. Determine the frequency of tweets and retweets featuring the hashtag #datasaveslives, including the most frequently shared tweets (how often was #datasaveslives tweeted?).
2. Characterize the range of stakeholder groups that use and share #datasaveslives and visualize retweet relationships between users (who tweets #datasaveslives and how were tweets shared between users?).
3. Identify and explore the different purposes that people used #datasaveslives for when tweeting (what did people use #datasaveslives to tweet about?).

Methods

Design and Objectives

We used a mixed methods design, combining elements of descriptive statistics, social network analysis, and qualitative research. This approach, which used a combination of qualitative and quantitative analysis, was adopted to allow a richer analysis of Twitter posts, over and above what could be achieved by available social media analytics tools.

Data Set, Variables, and Definitions

The data set comprised all publicly available tweets (N=13,895) that included the hashtag #datasaveslives posted between September 1, 2016, and August 31, 2017. This year was selected because it was perceived to represent a peak in campaign activity, thereby providing a sufficiently large and diverse sample of tweets for analysis. These were procured from Twitter's historical data service in January 2018.

The following variables pertaining to the tweet text and metadata associated with the tweet were retained for the analysis: tweet ID, tweet text (*body*), a list of hashtags included in tweets, number of retweets, and date posted (recoded into day, month, and year).

Twitter classifies each tweet as either an original *post* or a *share*. Posts were defined as tweets where the user either created a new tweet with their own original text or where a user shared another user's tweet and added new text to accompany it (*quote*

tweets). Shares (more commonly referred to as *retweets*) referred to cases when the user had shared a post created by another user with their followers without changing or adding new text. In all cases, tweets were only included if they referenced #datasaveslives somewhere in the body of the tweet, whether in the shared text or the text newly added by the user.

Where available, we also retained the following data pertaining to individual users who posted tweets, specifically: username, bio (optional self-written text about the user in 160 characters or less), friend count (users they had elected to follow), and follower count (users who had elected to follow them).

For analysis purposes, we defined official supporters as the 6 user accounts belonging to the sites of the Farr Institute and the Connected Health Cities (CHC) programme, all of whom adopted #datasaveslives as part of their formal strategies (@FarrInstitute, @CHCNorth, @HeRC_Farr, @FarrScotland, @FarrCIPHER, and @FarrLondon).

Data Preprocessing and Analysis

Historical Twitter data were preprocessed using Python (version 3.7.2). Briefly, the *pandas* Python library was used to convert data from a JavaScript Object Notation (JSON) format into a two-dimensional data frame for cleaning, recoding, and validation tasks in preparation for data analysis.

Statistical analyses were completed using RStudio (version 1.1.456). To address objective 1, descriptive statistics were used to determine weekly and monthly frequencies of tweets featuring #datasaveslives and percentages of the most commonly shared tweets (retweets). Pearson R was used to determine the associations between weekly counts for posts and retweets. For the most commonly shared retweets, the total potential reach was estimated by summing the follower count for every user who shared the tweet.

To address objective 2, users were grouped according to tweet frequency, and their characteristics were analyzed in terms of median counts for followers, friends, and posts. Gephi (version 0.9.2), a social network analysis tool, was used to analyze and visualize relationships between users of #datasaveslives. We focused on the retweet network as a way of understanding the sharing practices and underlying network structures between users. Statistics about the overall network and individual vertices were generated based on who retweeted whom, including clustering coefficients and measures of centrality. These were used to produce an undirected network graph visualizing the connections (edges) between users (vertices). To detect communities and calculate modularity, we used the Louvain method for community detection, which has been shown to outperform similar modularity methods in terms of speed and efficiency [20]. The graph was laid out using the Force Atlas layout algorithm. Common words used in user bios and tweet texts were also identified for each cluster (excluding commonly used words, eg, *and*, *or*, *views*).

To address objective 3, thematic analysis [21] was used to analyze the textual content of tweets featuring #datasaveslives qualitatively. Owing to the large size of the data set, it was not deemed practical or necessary to read and code all tweets. All original posts accompanying the hashtag were imported into NVivo 12 [22] for analysis. After reading a convenience sample of tweets (the first 200 tweets in date order), we defined an initial coding structure, covering the range of purposes tweets appeared to be used for. All original posts were sorted using the random number generator function in Microsoft Excel and then reviewed, coded, recoded, and collated into key themes in an iterative fashion by LH. Tweets were coded until saturation occurred, that is, until no substantially new themes were found. Approximately 1000 tweets were manually reviewed in total. The final set of themes was decided upon following a discussion between the authors.

Ethics and Governance

Data were collected and processed in line with Twitter's terms and conditions. As this information was nonsensitive and already in the public domain, formal ethical approvals were not required to complete the project.

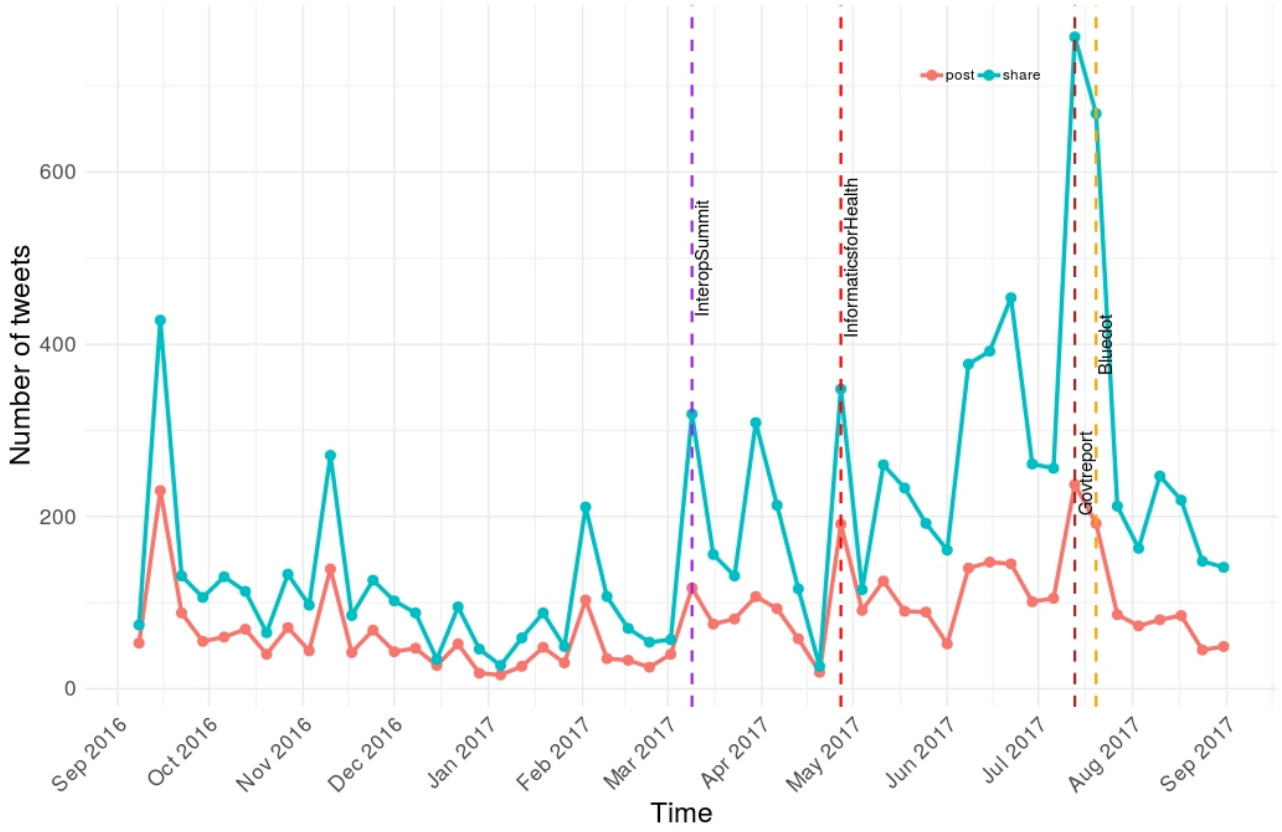
On advice from our university's research ethics office and in line with wider social media research guidelines [23], we took the following measures to protect Twitter users' privacy and confidentiality expectations: first, only tweets of users with accounts set to *public* were included in the analysis. Second, we gained permission to quote verbatim posts by individual users who were not clearly part of identifiable public groups or bodies or tweeting in an official capacity (eg, government organizations, university departments, heads of department). During the course of identifying popular tweets, we discovered that some posts or accounts had subsequently been deleted by users following the time of data collection; in such cases, tweets were not quoted although they were retained for the purposes of aggregated quantitative and qualitative analyses.

Results

How Often Was #datasaveslives Tweeted?

During the observation period, there were 13,895 tweets containing #datasaveslives (Figure 1). Overall, 30.05% (4175/13,895) were original posts and 69.95% (9720/13,895) were retweets. Among the original posts, 34.80% (1453/4175) were quote tweets. The mean number of total weekly tweets was 267.21 (SD 200.06), although this varied substantially (range 43-994). There was a strong positive correlation between weekly counts for posts and retweets (r=0.927, df=50; P<.001).

Figure 1. Tweet frequency over time by tweet type.



The highest number of tweets was observed during the week commencing July 6, 2017 (237 posts and 757 retweets), during which the UK government published a response [24] to a national review of security, consent processes, and opt-outs relevant to health data [25]. During the same week, there were also tweets about public engagement activities at high-profile cultural festivals in Cheshire (Bluedot Festival, England) and Edinburgh (Edinburgh Festival Fringe, Scotland). There were also high frequencies of tweets from official supporters during the week beginning April 20 (week 34), when there was a health informatics conference (Informatics for Health) hosted in Manchester in England.

Overall, 6 of the 10 most frequently shared tweets were from accounts associated with organizations, networks, or events (Table 1); only 1 originated from the account of an official supporter (@HerC_Farr). There was a modest, though significant, positive correlation between retweet count and follower count (r=0.214, df=4173; P<.001).

Table 1. The top 10 most frequently shared tweets.

Rank	Tweet[a]	Username and bio	Retweets, n	Group[b]	Total potential follower reach, n[c]
1	"Without data, this wouldn't be possible. We welcome the Govt's response to @NDGoffice review #DataSavesLives"	@NHSDigital; Information and technology for better health and care	58	7	274,311
2	"#DataSavesLives Our open letter from charities following the Government's response to the Caldicott Review"	@wellcometrust; We're a charitable foundation that exists to improve health for everyone. We support thousands of scientists & researchers, spark debate & take on big problems	51	3	206,171
3	Not available. Tweet deleted by the user	Bio not available	50	N/A[d]	N/A
4	"Remembering Alan Turing today, on his anniversary. An incredible scientist and human being, and an original believer in #datasaveslives"	@HeRC_Farr; An academic, NHS & Industry Partnership: Harnessing health data for patient and public benefit. #datasaveslives	41	1	52,544
5	"Better use of data means you don't have to tell your story again and again to doctors and nurses #DataSavesLives"	@NHSEngland; Health and high quality care for all, now and for future generations	40	10	212,478
6	"Using patient data is vital to improve health+care for us all #datasaveslives"	@NMRPerrin; Leading new Understanding Patient Data initiative. Interested in all things data, with a bit of science policy on the side	38	3	35,532
7	"Come + work with me! Understanding Patient Data team is recruiting a new policy/comms officer #datasaveslives"	@NMRPerrin; as above	37	3	138,791
8	"Register now for our Annual Scientific Meeting- Research in the Digital Age #DataSavesLives"	@SMHRN1; Scottish MH Research Network-supporting excellence in mental health studies as part of NHS Research Scotland	36	2	62,589
9	"New #INTEROPen board: an open collaboration of #interoperability networks to drive #OpenStandards in #health & #socialcare #DataSavesLives"	@INTEROPenAPI; Leading organizations supporting patients clinicians & new care models. Accelerating the delivery of #Interoperability #OpenStandards in health & social care	33	4	37,204
10	Not available. Tweet deleted by the user	Bio not available	31	N/A	N/A

[a]As of August 31, 2017.

[b]Group numbers cross-referenced with Table 3.

[c]Calculated as the sum of followers across all users who retweeted the original post. This method overestimates the total potential reach as it cannot account for the overlap of followers between users, and in any case, it is unlikely that all followers would view posts.

[d]N/A: not available.

Who Tweets #datasaveslives?

There were 3649 unique Twitter users who posted or shared content, including #datasaveslives (Table 2). Approximately 1 in 10 (1573/13895, 11.32%) of all #datasaveslives tweets, and 1 in 6 of posts that used original text (421/2722, 15.46%), were by official supporters. The tweet type was significantly associated with an official supporter status; official supporters used posts with original text relatively more often than others (26.76% vs 18.67%; χ_1^2=57.5; *P*<.001).

Among the 3649 users who posted or shared #datasaveslives at any time during the time window observed, 64.87% (2367/3649) did so only once (range 1-455). Users who tweeted 10 times or more accounted for just 4.88% (178/3649) of users, yet produced 54.33% (7549/13,895) of tweets; 16 users tweeted 100 times or more. This included 5 of the 6 official supporters, plus the accounts of affiliated organizations and projects. A total of 13 of the 16 accounts were associated with groups. In addition to official supporter organizations, these included health charities, professional membership organizations, event organizers, and projects. Notably, one of these frequent tweeters was a patient advocate and campaigner (n=102 tweets).

Table 2. Tweet frequency by tweet type and user type.

Tweet type	Tweet frequency by user type, n (%)		Total tweets (n=13,895), n (%)	Total unique users (n=3649), n (%)[a]
	Official supporter (n=1573)	Other (n=12,322)		
Original posts				
Original text	421 (26.76)	2301 (18.67)	2722 (19.60)	613 (16.80)
Quote	243 (15.44)	1210 (9.82)	1453 (10.46)	551 (15.10)
Shares				
Retweet	909 (57.79)	8811 (71.51)	9720 (69.95)	3157 (86.52)

[a]Owing to the overlap between users who use posts and shares, this column does not add up to 100%.

How Were Tweets Shared Between Users?

We visualized retweet relationships between Twitter users as an undirected network graph (Figure 2). Retweet connections were created when a user shared content by another user that included the hashtag. The analysis of retweets (n=9720) generated a network of 3392 users and 5749 unique connections between pairs of users (average degree=3.39; average path length=4.02; diameter=12).

Figure 2. Retweet network graph showing relationships between users who tweet and retweet #datasaveslives.

Cluster analysis using the Louvain method of community detection revealed 98 relatively well-connected groups (modularity=0.684). These were arranged in *hub and spoke* structures, with smaller numbers of relatively more tightly connected users at the center of each group. The 5 largest clusters or groups contained 60.70% (2059/3392) of users in

the network; 69 groups were very small, containing 5 or fewer users.

We examined the size, users, and words used in user bios for the largest 10 groups yielded by the cluster analysis (Table 3). The largest 2 groups (1 and 2) included all 6 official supporter accounts and were closely connected. Groups 1 and 2 shared similar vocabulary, both for tweets and user bios (eg, *health*, *research*, and *university*).

Table 3. Users in the #datasaveslives retweet user network by group.

Group	Users, n (%)	Most influential organizational user accounts (eigenvector centrality)[a]	Top 5 words used in user bios (n)	Top 5 words used in user tweets (n)
1	533 (15.71)	@CHCNorth (0.99), @HeRC_Farr (0.82), and The _NHSA (0.35)	Health (154), Manchester (71), research (67), university (64), and science (43)	Data (198), health (189), great (80), #iforh2017 (70) and research (53)
2	405 (11.94)	@FarrInstitute (1.0), @FarrScotland (0.79), and @FarrCIPHER (0.37)	Health (163), research (154), data (59), university (40), and public (39)	Data (143), health (96), research (51), #iforh2017 (46), and case (37)
3	399 (11.76)	@Patient_Data (0.43), @AMRC (0.36), and @wellcometrust (0.20)	Research (112), health (98), views (82), policy (44), and care (38)	Data (145), health (122), patient (60), information (57), and using (50).
4	390 (11.50)	@InteropSummit (0.33), @INTEROPenAPI (0.33), and @oht_uk (0.14)	Health (100), views (77), care (72), healthcare (44), and NHS[b] (40)	#interopsummit (110), #interoperability (53), #interopwarrior (38), care (36), and data (35)
5	332 (9.79)	@cancerchallscot (0.37), @IHDPscot (0.30), and @ProductForge (0.29)	Health (60), Scotland (51), care (38), data (37), and cancer (36)	#cancerdatadive (73), cancer (61), data (58), great (29), and #hackathon (27)
6	181 (5.34)	@GreatNorthCare (0.50) and @AHSN_NENC (0.19)	NHS (34), health (33), clinical (22), care (21), and director (17)	NHS (34), health (33), views (32), clinical (22), and care (21)
7	151 (4.45)	@NHSDigital (0.31), @DeptHealthPress (0.16), and @Soc_Endo (0.13)	Health (26), digital (22), care (15), research (14), and NHS (13)	Views (36), health (26), digital (22), care (15), and research (14)
8	127 (3.74)	@useMYdata (0.32), @DNADigest (0.10), and @abcdiagnosis (0.09)	Cancer (54), breast (29), research (27), health (18), and advocate (13)	Cancer (54), breast (29), research (27), health (18), and views (17)
9	117 (3.45)	@UoLCardioEpi (0.18), @LabKey (0.04), and @HealthSciYork (0.02)	Research (24), care (4), health (14), cardiovascular (13), and university (10)	Research (24), care (14), health (14), cardiovascular (13), and views (13)
10	110 (3.24)	@NHSEngland (0.17), @MedineGov (0.10), and @CURE_ScHARR (0.04)	Health (23), care (16), views (13), NHS (12), research (9), and healthcare (9)	Health (23), care (16), views (13), NHS (12), and research (9)

[a]Maximum of 3 users in the top 10 accounts.

[b]NHS: National Health Service.

Closer examination indicated some distinctions between groups 1 and 2. Group 1 users were more strongly affiliated with Northern England (particularly Manchester), whereas group 2 users frequently referenced places, organizations, and events located in Scotland. Group 1 was closely connected with group 5, which had a distinct topic focus on cancer data. Group 2 showed a stronger connection with group 6, which was associated with major medical records information technology (IT) projects based in the North East of England. Group 6 was, in turn, connected with group 4, populated by National Health Service (NHS) staff and delegates of a major health care IT conference (indicated by *#interopsummit*).

Group 3 indicated connections with both groups 1 and 2, and included users with connections to the NHS, health care policy, and major charities. Commonly used words in this group suggested a more applied focus among users (eg, *policy* and *care*). Group 3 was also loosely connected to group 8, distinctly notable for comprising users who self-identified as patients, carers, and advocates.

What Did People Use #datasaveslives to Tweet About?

The thematic analysis of tweet content yielded 4 key ways in which #datasaveslives was used: to share information and updates, for reporting and discussion at events, to show support for data sharing, and as a call to action. Although themes have been described separately for clarity, in practice there was substantial overlap, with the same tweets often being classified under multiple themes (Table 4).

Table 4. Examples of tweets with overlapping themes.

Theme[a]				Example tweets
A	B	C	D	
✓[b]	✓	N/A[c]	N/A	"Today is #WorldHealthDay - Find out how we work to improve health & care for patients & public here: [link to website] #datasaveslives" [@FarrInstitute]
✓	N/A	✓	N/A	"Interesting paper from @[usernames] calls for clarity on conflicting data sharing guidance [link to website] #datasaveslives" [@Patient_Data]
✓	N/A	N/A	N/A	"We are using patient data to implement learning health systems across the #North. Find out more: [link to website] #datasaveslives" [@AMRC]
N/A	✓	✓	N/A	"The Farr Institute discusses importance of patient data at House of Commons event #APPGMedResearch #datasaveslives [link to website]" [@FarrInstitute]
N/A	✓	N/A	N/A	"Thank you to all of our speakers today, to find out more about their work follow @UoLCardioEpi #datasaveslives #LIDASeminar" [@LIDA_UK]
N/A	N/A	✓	✓	"Everybody should be able to find out how patient data is used. Read our case studies on how #datasaves-lives… [link to website]" [@Patient_Data]
✓	✓	✓	N/A	"We believe #DataSavesLives! As do #interopsummit lecturers VIDEOS of Day 2 lectures on @YouTube [link to website] #interoperability" [@InteropSummit]
N/A	✓	✓	✓	"If you're at #IforH2017 don't forget to take a selfie with #datasaveslives at our stall (12) - just like [first name] from @[username] [photo]" [@FarrScotland]
✓	N/A	✓	✓	"Help contribute to the latest inquiry by @LordsSTCom into the #LifeSciences #IndustrialStrategy and highlight that #datasaveslives [link to website]" [@AMRC]

[a]Qualitative theme descriptions: A, to index and share information; B, for reporting and discussion at events; C, to show support for data sharing; and D, as a call to action.

[b]Data are applicable to themes.

[c]N/A: not applicable.

To Index and Share Information

The most common types of posts featuring #datasaveslives, particularly by official supporters and members of groups 1, 2, and 3 (Figure 2 and Table 3), were tweets sharing information about users' own projects, research findings, and news. These included announcements about new projects or funding, updates on progress, and sharing results from research. Although some tweets directly referenced peer-reviewed scientific literature by linking to journal publications, more often they were linked to less formal sources, including project websites, case studies, blogs, and videos:

> *Thanks to data we know that the smoking ban in Scotland has been a success [link to case study on website] #datasaveslives* [@FarrScotland, Group 2]

> *Highlights from Informatics for Health 2017 by @HeRC_Farr: Watch the video at [website link] #IforH2017 #datasaveslives* [@FarrInstitute, Group 2]

Twitter users also used #datasaveslives to highlight the work of others and signpost wider news and policy developments in areas relevant to health data science. These included news stories published by universities, health service organizations, professional bodies, and reports in popular media, including the local and national press and television and radio programs:

> *BBC News - Artificial intelligence predicts when heart will fail [link to news report] #DataSavesLives* [@EmpowerD4H, Group 13]

In the vast majority of cases, references for data sharing were positive or at least neutral; occasionally, however, there was evidence of more critical commentary about certain uses of health data:

> *Check out how @ukhomeoffice using health information is denying patients healthcare [link to news story] #DataSavesLives until it doesn't* [@einsteinsattic, Group 2]

Among tweets in this category, hyperlinks to other websites were very common; indeed, a subgroup of tweets were identified that included a hyperlink and the hashtag, indicating the use of #datasaveslives as purely an index function. This was mainly used by official supporters.

For Reporting and Discussion at Events

Frequently, #datasaveslives was used to tag tweets related to events, including conferences, meetings, and public engagement activities. Tweets included the promotion of forthcoming events, discussion of past events, or even live reporting and commentary about events, talks, and discussions that were currently underway. In the case of larger events, such as conferences, #datasaveslives frequently appeared alongside other official event hashtags (eg, #iforh2017, #interopsummit). Images of slides, presenters, delegates, visitors, and stalls were commonly included alongside the text:

> *Looking forward to meetings workshops and exciting stuff at @ExpoNHS tmrw #datasaveslives#nhs* [@ruthlady, Group 1]

To Show Support for Data Sharing

One further use of #datasaveslives was to demonstrate personal support for sharing health data in general or backing the #datasaveslives campaign itself. A total of 26 users included the text #datasaveslives within their Twitter bio. Many tweets of this type included images of individuals or groups at events pointedly posing with eye-catching placards, badges, or clothing featuring the hashtag:

> *Thanks for coming to chat wear your badge with pride!* [@FarrInstitute, Group 1]

Some tweets included a positive statement about reasons for supporting data sharing, either within the tweet or written on placards pictured in the tweet. The reasons referenced included sharing health data for research, sharing data as part of routine health care, or sharing data as part of larger projects that combined elements of both. Some tweets within this category signposted wider evidence supporting data sharing, such as collections of case studies where health data had been used for patient benefit. These were especially common among groups 4 and 5. Some drew on first-hand experiences and opinions:

> *For more examples of how #datasaveslives in mental health read this @MQmentalhealth blog. See our case studies [link to website]* [@Patient_Data, Group 3]

> *The type of treatment that I had depended so much on the data of patients who went before me' - patient advocate - #datasaveslives* [@useMYdata, Group 8]

As a Call to Action

We also identified a category of tweets that were used to make requests for others to act, participate, or respond in some manner. Commonly, these included advertisements to register for or submit papers to future events, participate in research studies, visit exhibition stands at conferences, or apply for jobs. There were also requests to provide feedback, opinions, or information:

> *We're inviting applications for a 2yr Clinical Research Fellow to study for an MD. Cardiology trainees please. #heartattack #datasaveslives* [@UoLCardioEpi, Group 9]

> *Help guide our consent modelling framework: happy to share a copy of your care org's consent forms? TY/please DM #datasaveslives #ontology* [@GreatNorthCare, Group 6]

Discussion

Principal Findings

This study investigated how a dedicated hashtag was used to promote the reuse of health data for research purposes and public benefit, how often, and by whom. Originally launched by the Farr Institute for Health Informatics Research, #datasaveslives came to be adopted by several distinct, diverse, yet interconnected groups in the United Kingdom with shared interests in health informatics, policy, and research. Our findings suggest that reasons for tweeting #datasaveslives evolved beyond the original objective of indexing information to a broader range of purposes, including event reporting, encouraging participation and action, and showing support for sharing health data.

Comparisons With Previous Work

Among the wider range of communities who shared content tagged with #datasaveslives, we detected 2 communities in particular who were research-focused, geographically distinct, and strongly interconnected. These were, in turn, connected with distinct professional communities with wider interests—some with access to sizable networks, funding, and influence—including government departments, the NHS, policy makers, patient advocates, and major charities. Our findings fit with the wider literature, which indicates that scientists can use Twitter not only to communicate with each other but also to engage broader audiences, including policy makers and the public [3,6].

One of the initial, more obvious uses of the hashtag was to index information about the use of health data as part of research and innovation, and make it more readily retrievable to a wider, not exclusively scientific, audience. Moreover, people also used the hashtag to publicly demonstrate support for data sharing and each other. This is compatible with the wider literature, which suggests that academics use hashtags to categorize information [26] and encourage interaction and community building [27-29]. These uses seem pertinent, given that our period of observation followed the high-profile failure of the care.data scheme, a major government initiative in England to share patient data [30]. Indeed, two of the most frequently shared tweets in our analysis concerned subsequent proposals to change government policy, addressing data security and consent [24,25]. Previous studies have shown how responses to care.data on Twitter attracted critical commentary [31], including from interest communities in politics, health care, and the media [32]. Before the observation period examined in this study, concerns had been raised about access to patient data by commercial companies, especially where these uses were perceived to be primarily motivated by profit rather than public benefit [30,33-35]. This study contrasts with these findings, showing how #datasaveslives was used in the wake of public backlash to care.data to spread mainly positive messages about data use and reuse, and to increase transparency, demonstrate solidarity, and provide supportive networks among health, data, and IT professionals.

In declaring an intent to promote the reuse of health data for research purposes, the #datasaveslives campaign could be regarded as a behavioral intervention of sorts, encouraging credible users to endorse and share supportive messages. As with other behavioral interventions conducted via social media, attention should be directed toward identifying the active ingredients of interventions [36]. Our thematic analyses of tweet content revealed 2 noteworthy and interrelated strategies used as components to achieve campaign aims. First, #datasaveslives was used at events frequented by influential communities, generating spikes in activity generated by commentary about the proceedings of meetings and events in real time. So-called *live-tweeting* has become more common at scientific conferences and has the advantage of increasing transparency and rapidly

disseminating information among a far larger audience over and above those who physically attend [37,38]. Using #datasaveslives, either alone or in addition to more specific conference hashtags, might have amplified the reach of information while avoiding the limited audience and *shelf-life* of more niche conference hashtags.

Second, offline activities at events were used to drive the generation of web-based multimedia content; events were used as photo opportunities for individuals willing to publicly endorse #datasaveslives, leveraged by attractive branded physical merchandise. Drawing on evidence from previous studies, which have identified health behavior change techniques particular to social media, reviewed by Simeon et al [36], these photo opportunities might be framed as *virtual rewards*, in turn encouraging further *overt endorsements* in the form of likes, retweets, and comments. Indeed, similar social media strategies have been used in both the health sciences and the corporate sectors, such as identifying target communities, gaining support from credible and/or influential users, developing engaging multimedia content, updating content regularly, improving the visual presentation of content, and encouraging participation via small concrete actions [16,39,40].

Strengths and Limitations

This study benefits from the analysis of a near-complete sample of #datasaveslives public tweets for an active year during the campaign. Nonetheless, we could not have captured all mentions and uses of #datasaveslives during this period. Private and previously deleted tweets were excluded. Owing to the limited use of other social media platforms by official supporters, our analysis only considered Twitter posts tagged with the keyword #datasaveslives. It is notable that other important public health outreach campaigns—including during outbreaks [41], as part of science communication [16] and to promote health behavior change [36]—have commonly used a wider range of social media platforms, particularly Facebook. The content, strategies, and communities observed in this study may be specific to Twitter and should not be generalized to other social media or content-sharing platforms. Furthermore, the network analysis was limited to retweets; we did not capture other types of engagement, such as follower networks, or use directed networks, as done by other studies [32]. Thus, certain nuances of information flow may have been lost, indicating influential relationships. Demographic data about users were not made available by Twitter for analysis, limiting our understanding of sample characteristics. Finally, we accept that we were unable to quantify, much less characterize, the much wider audience who saw, read, or otherwise engaged with tweets, in particular patients and members of the wider public not connected to organizations.

Future Research

The health data science community has stated a vision to be team-based, transparent, and inclusive, seeking involvement from a wide range of interdisciplinary stakeholders, including patients and the public [42]. Future research would benefit from examining how the use and users of #datasaveslives have changed over time and suitable ways of determining the overall impact of varying strategies to engage key communities, such as members of the public. Using such opportunities for social media to contribute toward building networks and engaging in dialog in open forums would seem eminently compatible with this vision.

Conclusions

The rise of social media has provided unprecedented opportunities for academic organizations and individual scientists to communicate with a much wider range of stakeholders than ever before, including the public. This study shows how a simple hashtag campaign on Twitter was used to disseminate credible scientific information and increase the visibility of research activities, with evidence to suggest this supported community building and bridging practices among interdisciplinary sectors allied to health data science.

Our findings are of interest to a variety of stakeholders who share an interest in supporting the reuse of health data for public benefit. By revealing the different communities who share such interests, analyzing content thematically, and demonstrating how information flows between them, our findings can be used to better understand the mechanisms underpinning stakeholder engagement campaigns conducted on social media and how to optimize these further.

Acknowledgments

This paper is dedicated to the memory of Alan Campbell. LH is funded via a UK Research and Innovation Fellowship at Health Data Research UK (ref: MR/S004025/1). Access to the data was funded by CHC, which was funded by the Department of Health and Social Care in England. The views expressed are those of the authors and not necessarily those of the Health Data Research UK, CHC, or the Department of Health and Social Care. The authors thank James Cunningham and Miguel Belmonte for their assistance with data preparation and visualization. They also acknowledge the work of Stephen Melia, in tirelessly promoting #datasaveslives. He may not have been the first person to coin the phrase *data saves lives*, but he was (probably) the first to put a hashtag in front of it.

Conflicts of Interest

LH and MT were previously funded by the CHC program and the Health e-Research Center, which founded the #datasaveslives campaign reported in this paper.

References

1. Holmberg K, Thelwall M. Disciplinary differences in Twitter scholarly communication. Scientometrics 2014 Jan 22;101(2):1027-1042. [doi: 10.1007/s11192-014-1229-3]
2. Holmberg K, Bowman TD, Haustein S, Peters I. Astrophysicists' conversational connections on Twitter. PLoS One 2014;9(8):e106086 [FREE Full text] [doi: 10.1371/journal.pone.0106086] [Medline: 25153196]
3. Priem J, Costello KL. How and why scholars cite on Twitter. Proc Am Soc Info Sci Tech 2011 Feb 3;47(1):1-4. [doi: 10.1002/meet.14504701201]
4. Wilkinson SE, Basto MY, Perovic G, Lawrentschuk N, Murphy DG. The social media revolution is changing the conference experience: analytics and trends from eight international meetings. BJU Int 2015 May;115(5):839-846. [doi: 10.1111/bju.12910] [Medline: 25130687]
5. Shiffman D. Twitter as a tool for conservation education and outreach: what scientific conferences can do to promote live-tweeting. J Environ Stud Sci 2012 Jul 24;2(3):257-262. [doi: 10.1007/s13412-012-0080-1]
6. Côté IM, Darling ES. Scientists on Twitter: preaching to the choir or singing from the rooftops? Facets 2018 Oct 1;3(1):682-694. [doi: 10.1139/facets-2018-0002]
7. Bonetta L. Should you be tweeting? Cell 2009 Oct 30;139(3):452-453 [FREE Full text] [doi: 10.1016/j.cell.2009.10.017] [Medline: 19879830]
8. van Noorden R. Online collaboration: scientists and the social network. Nature 2014 Aug 14;512(7513):126-129. [doi: 10.1038/512126a] [Medline: 25119221]
9. Collins K, Shiffman D, Rock J. How are scientists using social media in the workplace? PLoS One 2016;11(10):e0162680 [FREE Full text] [doi: 10.1371/journal.pone.0162680] [Medline: 27732598]
10. Hanrahan BV, Convertino G, Nelson L. Modeling Problem Difficulty and Expertise in Stackoverflow. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion. New York, NY, USA: Association for Computing Machinery; 2012 Feb Presented at: ACM'12; February 11-15, 2012; Seattle, Washington, USA p. 91-94. [doi: 10.1145/2141512.2141550]
11. Winandy M, Kostkova P, de Quincey E, St Louis C, Szomszor M. Follow #eHealth2011: measuring the role and effectiveness of online and social media in increasing the outreach of a scientific conference. J Med Internet Res 2016 Jul 19;18(7):e191 [FREE Full text] [doi: 10.2196/jmir.4480] [Medline: 27436012]
12. Li J, Greenhow C. Scholars and social media: tweeting in the conference backchannel for professional learning. Educ Media Int 2015 Feb 16;52(1):1-14. [doi: 10.1080/09523987.2015.1005426]
13. Parra D, Trattner C, Gómez D, Hurtado M, Wen X, Lin Y. Twitter in academic events: a study of temporal usage, communication, sentimental and topical patterns in 16 computer science conferences. Comput Commun 2016 Jan;73:301-314. [doi: 10.1016/j.comcom.2015.07.001]
14. What Works Engaging the Public Through Social Media. National Co-ordinating Centre for Public Engagement. 2018. URL: https://www.publicengagement.ac.uk/sites/default/files/publication/what_works_engaging_the_public_through_social_media_november_2018.pdf [accessed 2021-01-22]
15. Choo E, Ranney M, Chan T, Trueger N, Walsh A, Tegtmeyer K, et al. Twitter as a tool for communication and knowledge exchange in academic medicine: a guide for skeptics and novices. Med Teach 2014 Dec 19;37(5):411-416. [doi: 10.3109/0142159x.2014.993371]
16. Fontaine G, Maheu-Cadotte M, Lavallée A, Mailhot T, Rouleau G, Bouix-Picasso J, et al. Communicating science in the digital and social media ecosystem: scoping review and typology of strategies used by health scientists. JMIR Public Health Surveill 2019 Sep 3;5(3):e14447 [FREE Full text] [doi: 10.2196/14447] [Medline: 31482854]
17. Data Saves Lives. URL: https://www.datasaveslives.info/ [accessed 2021-01-22]
18. Hemingway H, Lyons R, Li Q, Buchan I, Ainsworth J, Pell J, et al. A national initiative in data science for health: an evaluation of the UK Farr Institute. Int J Popul Data Sci 2020 Apr 8;5(1):1128 [FREE Full text] [doi: 10.23889/ijpds.v5i1.1128] [Medline: 32935051]
19. Twitter. URL: https://twitter.com [accessed 2021-01-22]
20. Blondel VD, Guillaume J, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech 2008 Oct 9;2008(10):P10008. [doi: 10.1088/1742-5468/2008/10/p10008]
21. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
22. Nvivo Qualitative Data Analysis Software. QSR International Pty Ltd. 2018. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2021-01-22]
23. Soucek P, Cihelkova I. Photodynamic therapy with Verteporfin in subfoveal choroidal metastasis of breast carcinoma (a controlled case). Neuro Endocrinol Lett 2006 Dec;27(6):725-728. [Medline: 17187011]
24. Department of Health, Data Sharing and Cyber Security Team. Your Data: Better Security, Better Choice, Better Care. Department of Health. 2017 Jul 12. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/627493/Your_data_better_security_better_choice_better_care_government_response.pdf [accessed 2021-01-22]

25. National Data Guardian for Health and Care: Review of Data Security, Consent and Opt-Outs. The National Data Guardian. 2016 Jun. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF [accessed 2021-01-22]
26. Archibald M, Clark A. Twitter and nursing research: how diffusion of innovation theory can help uptake. J Adv Nurs 2014 Jan 23;70(3):e3-e5. [doi: 10.1111/jan.12343]
27. Su L, Scheufele D, Bell L, Brossard D, Xenos M. Information-sharing and community-building: exploring the use of Twitter in science public relations. Sci Commun 2017 Oct 4;39(5):569-597. [doi: 10.1177/1075547017734226]
28. Lovejoy K, Saxton G. Information, community, and action: how nonprofit organizations use social media. J Comput Commun 2012;17(3):337-353. [doi: 10.1111/j.1083-6101.2012.01576.x]
29. Liu X, He Q, Tian Y, Lee W. Event-based Social Networks: Linking the Online and Offline Social Worlds. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.: Association for Computing Machinery; 2012 Presented at: KDD'12; August 12-16, 2012; Beijing, China p. 1032-1040 URL: https://dl.acm.org/citation.cfm?id=2339693 [doi: 10.1145/2339530.2339693]
30. Carter P, Laurie G, Dixon-Woods M. The social licence for research: why care.data ran into trouble. J Med Ethics 2015 May;41(5):404-409 [FREE Full text] [doi: 10.1136/medethics-2014-102374] [Medline: 25617016]
31. Hays R, Daker-White G. The care.data consensus? A qualitative analysis of opinions expressed on Twitter. BMC Public Health 2015 Sep 2;15:838 [FREE Full text] [doi: 10.1186/s12889-015-2180-9] [Medline: 26329489]
32. Amor B, Vuik S, Callahan R, Darzi A, Yaliraki S, Barahona M. Community detection and role identification in directed networks: understanding the Twitter network of the care.data debate. In: Adams N, Heard N, editors. Dynamic Networks and Cyber-Security. London, UK: World Scientific; Mar 22, 2016:111-136.
33. Pollock A, Roderick P. Trust in the time of markets: protecting patient information. The Lancet 2014 May;383(9928):1523-1524. [doi: 10.1016/s0140-6736(14)60727-3]
34. Ipsos MORI, Social Reserch Institute. The One-way Mirror: Public Attitudes To Commercial Access to Health Data. Wellcome Trust. 2016 Mar. URL: https://www.ipsos.com/ipsos-mori/en-uk/commercial-access-health-data [accessed 2021-01-22]
35. Tully MP, Hassan L, Oswald M, Ainsworth J. Commercial use of health data-a public 'trial' by citizens' jury. Learn Health Syst 2019 Oct;3(4):e10200 [FREE Full text] [doi: 10.1002/lrh2.10200] [Medline: 31641688]
36. Simeon R, Dewidar O, Trawin J, Duench S, Manson H, Pardo JP, et al. Behavior change techniques included in reports of social media interventions for promoting health behaviors in adults: content analysis within a systematic review. J Med Internet Res 2020 Jun 11;22(6):e16002 [FREE Full text] [doi: 10.2196/16002] [Medline: 32525482]
37. Ekins S, Perlstein EO. Ten simple rules of live tweeting at scientific conferences. PLoS Comput Biol 2014 Aug;10(8):e1003789 [FREE Full text] [doi: 10.1371/journal.pcbi.1003789] [Medline: 25144683]
38. Lister AL, Datta RS, Hofmann O, Krause R, Kuhn M, Roth B, et al. Live coverage of scientific conferences using web technologies. PLoS Comput Biol 2010 Jan 29;6(1):e1000563 [FREE Full text] [doi: 10.1371/journal.pcbi.1000563] [Medline: 20126525]
39. Freeman B, Potente S, Rock V, McIver J. Social media campaigns that make a difference: what can public health learn from the corporate sector and other social change marketers? Public Health Res Pract 2015 Mar 30;25(2):e2521517 [FREE Full text] [doi: 10.17061/phrp2521517] [Medline: 25848735]
40. Edney S, Bogomolova S, Ryan J, Olds T, Sanders I, Maher C. Creating engaging health promotion campaigns on social media: observations and lessons from fitbit and garmin. J Med Internet Res 2018 Dec 10;20(12):e10911 [FREE Full text] [doi: 10.2196/10911] [Medline: 30530449]
41. Raamkumar AS, Tan SG, Wee HL. Measuring the outreach efforts of public health authorities and the public response on Facebook during the COVID-19 pandemic in early 2020: cross-country comparison. J Med Internet Res 2020 May 19;22(5):e19334 [FREE Full text] [doi: 10.2196/19334] [Medline: 32401219]
42. Ford E, Boyd A, Bowles JK, Havard A, Aldridge RW, Curcin V, et al. Our data, our society, our health: a vision for inclusive and transparent health data science in the United Kingdom and beyond. Learn Health Syst 2019 Jul;3(3):e10191 [FREE Full text] [doi: 10.1002/lrh2.10191] [Medline: 31317072]

Abbreviations

CHC: Connected Health Cities
IT: information technology
NHS: National Health Service

Edited by G Eysenbach, R Kukafka; submitted 20.09.19; peer-reviewed by Z Ma, A Cyr, MS Aslam, K Goniewicz; comments to author 18.05.20; revised version received 27.07.20; accepted 07.12.20; published 16.02.21.

Please cite as:
Hassan L, Nenadic G, Tully MP
A Social Media Campaign (#datasaveslives) to Promote the Benefits of Using Health Data for Research Purposes: Mixed Methods Analysis
J Med Internet Res 2021;23(2):e16348
URL: http://www.jmir.org/2021/2/e16348/
doi:10.2196/16348
PMID:33591280

Letter to the Editor

Data Leakage in Health Outcomes Prediction With Machine Learning. Comment on "Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning"

Alexandre Chiavegatto Filho[1], PhD; André Filipe De Moraes Batista[1], MSc, PhD; Hellen Geremias dos Santos[1], MPH, PhD

Department of Epidemiology, School of Public Health, University of São Paulo, São Paulo, Brazil

Corresponding Author:
Alexandre Chiavegatto Filho, PhD
Department of Epidemiology
School of Public Health
University of São Paulo
Av Dr Arnaldo 715
São Paulo
Brazil
Phone: 55 955543047
Email: alexdiasporto@usp.br

Related Article:

Comment on: https://www.jmir.org/2018/1/e22/

(J Med Internet Res 2021;23(2):e10969) doi:10.2196/10969

KEYWORDS

machine learning; data leakage; prediction

Applications of machine learning algorithms to predict the incidence of health outcomes have an enormous potential to improve clinical practice and lower health care costs [1]. Machine learning is a subset of artificial intelligence that uses data to improve decisions through experience, which is especially promising in a data-driven world. Dr Ye and colleagues' article on hypertension incidence prediction in the *Journal of Medical Internet Research* adds to this literature [2], but its potential contribution and applicability are hindered by a major flaw.

The objective of the study was to "develop and validate prospectively a risk prediction model of incident essential hypertension within the following year." The authors follow good prediction protocols by applying a high-performing machine learning algorithm (XGBoost) and by validating the results on unseen data from the following year. The algorithm attained a very high area under the curve (AUC) value of 0.870 for incidence prediction of hypertension in the following year.

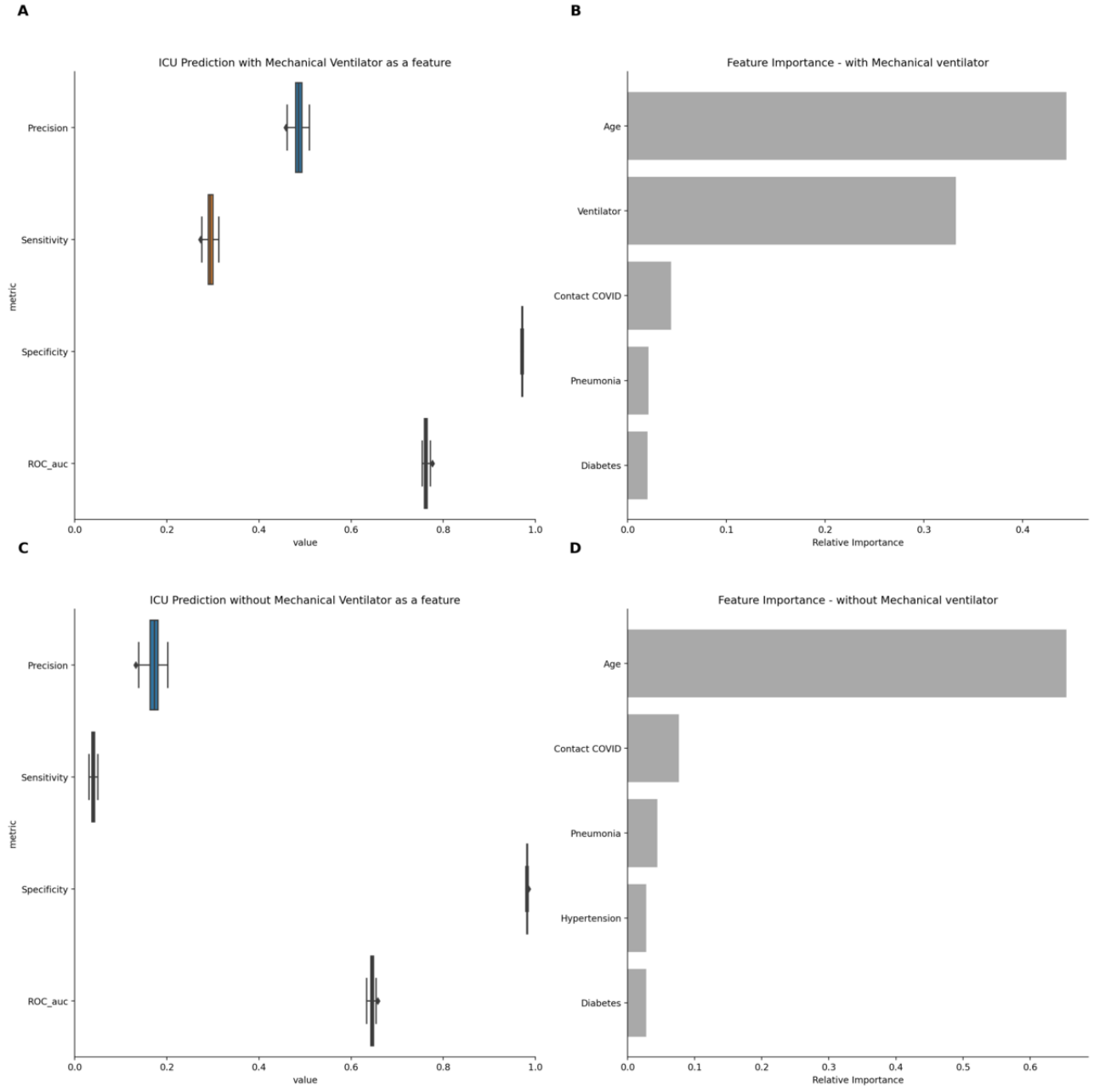
The authors follow this impressive result by commenting on some of the most important predictive variables, such as demographic features, diagnosed chronic diseases, and mental illness. The ranking of the variables that were most important for the predictive performance of hypertension is included in a multimedia appendix; however, the above-mentioned variables are not listed near the top. Of the six most important variables, five were: lisinopril, hydrochlorothiazide, enalapril maleate, amlodipine besylate, and losartan potassium. All of these are popular antihypertensive drugs.

Data leakage occurs when one or more features used to train the algorithm has hidden within itself the result of the outcome, and is considered one of the most frequent mistakes in machine learning [3]. This is different from predictive importance, that is, the relative effect of each variable in increasing or decreasing the expected outcome, as it usually comes after the outcome. Therefore, it is a consequence of the outcome that is being predicted and not the other way around.

A classic example from machine learning textbooks is the inclusion of the ID number of the patient as a predictor. While this should not have predictive importance if randomly assigned, it is common that patients coming from the same hospital have similar ID numbers in multicenter data sets. In the case of cancer prediction, for example, machine learning algorithms will learn that similar ID numbers that come from oncology hospitals have a higher probability of cancer.

As an example, we used real data to test the effect of including mechanical ventilation to predict intensive care unit (ICU) admission among patients with COVID-19 [4]. This is another example of data leakage, as mechanical ventilation usually only occurs after ICU admission and should not be used to predict its risk. Figure 1 shows the decrease in the prediction metrics for ICU admission with the exclusion of mechanical ventilation as a predictor, with the area under the ROC (receiver operating characteristic) curve decreasing from 0.76 to 0.64, and precision from 0.49 to 0.17.

Figure 1. Performance metrics for the prediction of intensive care unit (ICU) admission with and without the use of mechanical ventilation as a predictor.



By including the use of antihypertensive drugs as predictors for hypertension incidence in the following year, Dr Ye and colleagues' work opens the possibility that the machine learning algorithm will focus on predicting those already with hypertension but did not have this information on their medical record at baseline. While this would work for a prediction competition, where data science teams compete to produce the best predictive model such as in a Kaggle challenge [5], it is not of particular scientific or clinical interest. In the case of the latter, just one variable (the use of a hypertension drug) is sufficient for physicians to infer the presence of hypertension, while for the former, the knowledge of this being a highly predictable event (as measured by the AUC) is severely impaired.

In order to identify the presence of data leakage in prediction studies, it is important to have a conceptual pathway of how the predictors longitudinally affect the outcome variable, as there is no statistical method that is capable of pointing out the presence of data leakage. Improving the predictive performance of specific data sets for different diseases is an important new field in epidemiology and data science. The authors can still

contribute to this literature by providing the new AUC of the prediction after addressing the data leakage issue.

Editorial Notice

The corresponding author of "Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning" did not respond to our invitation to reply to this commentary.

Conflicts of Interest

None declared.

References

1. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016 Sep 29;375(13):1216-1219 [FREE Full text] [doi: 10.1056/NEJMp1606181] [Medline: 27682033]
2. Ye C, Fu T, Hao S, Zhang Y, Wang O, Jin B, et al. Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning. J Med Internet Res 2018 Jan 30;20(1):e22 [FREE Full text] [doi: 10.2196/jmir.9268] [Medline: 29382633]
3. Kaufman S, Rosset S, Perlich C. Leakage in Data Mining: Formulation, Detection, and Avoidance. 2011 Presented at: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 21-24, 2011; San Diego, CA, USA p. 556-563. [doi: 10.1145/2020408.2020496]
4. Database: Covid-mexico (predictive models). GitHub. 2020. URL: https://github.com/salomonw/covid-mexico [accessed 2020-11-01]
5. Nisbet R, Elder J, Miner G. Handbook of Statistical Analysis and Data Mining Applications. Burlington, MA: Elsevier; 2009.

Abbreviations

AUC: area under the curve
ICU: intensive care unit
ROC: receiver operating characteristic

Edited by G Eysenbach, T Derrick; submitted 04.05.18; peer-reviewed by J Goris; comments to author 25.10.20; revised version received 08.11.20; accepted 27.01.21; published 11.02.21.

Please cite as:
Chiavegatto Filho A, Batista AFDM, dos Santos HG
Data Leakage in Health Outcomes Prediction With Machine Learning. Comment on "Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning"
J Med Internet Res 2021;23(2):e10969
URL: https://www.jmir.org/2021/2/e10969
doi:10.2196/10969
PMID:33570496

Letter to the Editor

The Potential for the Internet and Telehealth in Caregiver Support. Comment on "Using Technology to Facilitate Fidelity Assessments: The Tele-STAR Caregiver Intervention"

Kohei Kajiwara[1], PhD; Jun Kako[2], PhD; Hiroko Noto[3], PhD; Yasufumi Oosono[4], PhD; Masamitsu Kobayashi[4], MSN

[1]Japanese Red Cross Kyushu International College of Nursing, Munakata, Japan

[2]College of Nursing Art and Science, University of Hyogo, Akashi, Japan

[3]Department of Health Sciences, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

[4]Faculty of Nursing, National Defense Medical College, Tokorozawa, Japan

Corresponding Author:
Kohei Kajiwara, PhD
Japanese Red Cross Kyushu International College of Nursing
1-1 Asty Munakata-City
Fukuoka-Prefecture 811-4157
Japan
Phone: 81 940 35 7030
Email: k-kajiwara@jrckicn.ac.jp

Related Article:

Comment on: https://www.jmir.org/2019/5/e13599/

(J Med Internet Res 2021;23(2):e14953) doi:10.2196/14953

KEYWORDS

dementia; caregiver; technology

We read with interest the recent paper entitled "Using Technology to Facilitate Fidelity Assessments: The Tele-STAR Caregiver Intervention" by Lindauer et al [1]. The authors concluded that Tele-STAR contributed to low caregiver burden and showed good fidelity as an intervention method.

Internet-based videoconferencing technology is an important source of support for caregivers of persons with dementia. Researchers have previously demonstrated the positive potential of computer-mediated interventions and technology-based cognitive behavioral therapy interventions for caregivers of people with dementia [2,3]. Others have raised the difficulties in measuring intervention fidelity in a consistent manner [4,5], which raises the importance of consistency when considering fidelity evaluations. Moreover, as the study reported results that may be attributed to both in-home and telehealth intervention experiences of participants, it may be useful to consider the interplay of these aspects.

Lindauer and colleagues [1] reported a slight reduction in caregiver burden, attributed to an improvement in caregivers' responses to patients with dementia, facilitated by the Tele-STAR intervention. Caregiver burden is an important consideration in the field of dementia care. A recent study found an internet-based intervention to be effective in increasing the positive aspect of subjective appraisal for caregivers of persons with dementia [6]. In addition, we have studied the subjective appraisal of both negative and positive aspects in this population [7]. Assessments that take into account both sides of subjective appraisal are capable of providing a broad understanding of a caregiver's context, and we would argue that an outcome that takes both into account would be more useful than current practices allow for.

The support that can be offered to caregivers using internet-driven technologies should continue to be explored, and the study conducted by Lindauer and colleagues [1] provides useful data in this regard. We agree that internet-based interventions will be beneficial to caregivers of persons with dementia in the future.

Editorial Notice

The corresponding author of "Using Technology to Facilitate Fidelity Assessments: The Tele-STAR Caregiver Intervention" declined to respond to this letter.

Conflicts of Interest

None declared.

References

1. Lindauer A, McKenzie G, LaFazia D, McNeill L, Mincks K, Spoden N, et al. Using Technology to Facilitate Fidelity Assessments: The Tele-STAR Caregiver Intervention. J Med Internet Res 2019 May 24;21(5):e13599 [FREE Full text] [doi: 10.2196/13599] [Medline: 31127721]
2. McKechnie V, Barker C, Stott J. Effectiveness of computer-mediated interventions for informal carers of people with dementia-a systematic review. Int Psychogeriatr 2014 Oct;26(10):1619-1637 [FREE Full text] [doi: 10.1017/S1041610214001045] [Medline: 24989249]
3. Scott JL, Dawkins S, Quinn MG, Sanderson K, Elliott KE, Stirling C, et al. Caring for the carer: a systematic review of pure technology-based cognitive behavioral therapy (TB-CBT) interventions for dementia carers. Aging Ment Health 2016 Aug;20(8):793-803. [doi: 10.1080/13607863.2015.1040724] [Medline: 25978672]
4. Godwin KM, Mills WL, Anderson JA, Kunik ME. Technology-driven interventions for caregivers of persons with dementia: a systematic review. Am J Alzheimers Dis Other Demen 2013 May;28(3):216-222 [FREE Full text] [doi: 10.1177/1533317513481091] [Medline: 23528881]
5. Nguyen TA, Nguyen H, Pham T, Nguyen TH, Hinton L. A cluster randomized controlled trial to test the feasibility and preliminary effectiveness of a family dementia caregiver intervention in Vietnam: The REACH VN study protocol. Medicine (Baltimore) 2018 Oct;97(42):e12553 [FREE Full text] [doi: 10.1097/MD.0000000000012553] [Medline: 30334942]
6. Moskowitz JT, Cheung EO, Snowberg KE, Verstaen A, Merrilees J, Salsman JM, et al. Randomized controlled trial of a facilitated online positive emotion regulation intervention for dementia caregivers. Health Psychol 2019 May;38(5):391-402 [FREE Full text] [doi: 10.1037/hea0000680] [Medline: 31045422]
7. Kajiwara K, Noto H, Yamanaka M. Changes in caregiving appraisal among family caregivers of persons with dementia: A longitudinal study over 12 months. Psychogeriatrics 2018 Nov;18(6):460-467. [doi: 10.1111/psyg.12360] [Medline: 30066482]

Edited by G Eysenbach, T Derrick; submitted 06.06.19; peer-reviewed by A Lindauer; accepted 27.01.21; published 16.02.21.

Please cite as:
Kajiwara K, Kako J, Noto H, Oosono Y, Kobayashi M
The Potential for the Internet and Telehealth in Caregiver Support. Comment on "Using Technology to Facilitate Fidelity Assessments: The Tele-STAR Caregiver Intervention"
J Med Internet Res 2021;23(2):e14953
URL: https://www.jmir.org/2021/2/e14953
doi:10.2196/14953
PMID:33591281

Letter to the Editor

Remote Examination and Screening for Domestic Abuse. Comment on “Online Antenatal Care During the COVID-19 Pandemic: Opportunities and Challenges”

Hannah Lee Grimes[1*], BA; Ramnik Uppal[1*], BA

University of Cambridge Clinical School, Cambridge, United Kingdom

*all authors contributed equally

Corresponding Author:
Hannah Lee Grimes, BA
University of Cambridge Clinical School
Hills Road
Cambridge
United Kingdom
Phone: 44 01223 336700
Email: hlg42@cam.ac.uk

Related Article:

Comment on: https://www.jmir.org/2020/7/e19916/

(J Med Internet Res 2021;23(2):e23295) doi:10.2196/23295

KEYWORDS

spouse abuse; domestic abuse; apps; patient information; antenatal care; COVID-19

After reading the viewpoint by Wu et al [1] on opportunities for the delivery of antenatal care online during the COVID-19 pandemic, we would like to draw attention to a potential further aspect of this analysis. The authors surveyed 983 Chinese pregnant women and concluded that online antenatal care is likely to be an effective alternative for pregnant women, enabling them to access basic care without attending the hospital[1]. This is optimal with regards to minimizing the infection risk during the pandemic. However, it is important to remember that for a proportion of women, these antenatal visits provide a critical opportunity for physicians to screen for domestic abuse.

A cross-sectional study, also assessing a population of Chinese pregnant women, found that 15.62% of their subjects were identified as victims of domestic abuse [2]. Data released by the World Health Organization suggest that the number of women affected by domestic violence during pregnancy may be between 1% and 28% [3]. This proportion is by no means negligible.

In many countries, antenatal screening for domestic abuse when consulting with the patient is sporadic. One study of Canadian doctors found that 33% of professionals surveyed “never or only rarely” asked women about domestic abuse during antenatal appointments [4]. Provision of antenatal care online could result in even fewer women being able to confide in their doctor and receive much-needed assistance with respect to abusive situations at home. This might be due to women feeling less able to seek help through an online service compared to face-to-face consultations.

The COVID-19 pandemic has complicated this situation further. The London metropolitan police reported that since lockdowns were introduced in the United Kingdom, 380 more calls per week related to domestic abuse have been recorded [5]. Unpacking this data further, it is noteworthy that the majority of these reports of abuse were from third-party callers, perhaps due to the increased presence of neighbors [5]. This suggests that while there may be a potential surge in domestic abuse due to lockdown restrictions, this may not be associated with an increased probability of victims reporting the crime.

We suggest that to best address this gap in their proposal, Wu et al [1] should factor in how their proposed online antenatal care delivery system could incorporate adaptations to provide help to victims without compromising their safety. One possible method of achieving this could be to add a layer to any application developed through which women could find support numbers for relevant domestic abuse charities or alert their doctor to their need for a face-to-face appointment to discuss their home situation. It is important for this feature to be sufficiently blended into the app such that it would not alert suspicion from the perpetrator and thus compromise the safety of any woman seeking help.

Editorial Notice

The corresponding author of "Online Antenatal Care During the COVID-19 Pandemic: Opportunities and Challenges" did not respond to our invitation to reply to this commentary.

Conflicts of Interest

None declared.

References

1. Wu H, Sun W, Huang X, Yu S, Wang H, Bi X, et al. Online Antenatal Care During the COVID-19 Pandemic: Opportunities and Challenges. J Med Internet Res 2020 Jul 22;22(7):e19916 [FREE Full text] [doi: 10.2196/19916] [Medline: 32658860]
2. Zheng B, Zhu X, Hu Z, Zhou W, Yu Y, Yin S, et al. The prevalence of domestic violence and its association with family factors: a cross-sectional study among pregnant women in urban communities of Hengyang City, China. BMC Public Health 2020 May 05;20(1):620 [FREE Full text] [doi: 10.1186/s12889-020-08683-9] [Medline: 32370810]
3. Garcia-Moreno C, Jansen HA, Ellsberg M, Heise L, Watts CH. Prevalence of intimate partner violence: findings from the WHO multi-country study on women's health and domestic violence. The Lancet 2006 Oct;368(9543):1260-1269. [doi: 10.1016/S0140-6736(06)69523-8]
4. Long AJ, Golfar A, Olson DM. Screening in the Prenatal Period for Intimate Partner Violence and History of Abuse: A Survey of Edmonton Obstetrician/Gynaecologists. J Obstet Gynaecol Can 2019 Jan;41(1):38-45. [doi: 10.1016/j.jogc.2018.05.003] [Medline: 30585166]
5. Ivandic R, Kirchmaier T. Home is not a safe place for everyone: domestic abuse between partners increased during lockdown. LSE Covid 19 Blog. 2020 Jun 30. URL: https://blogs.lse.ac.uk/covid19/2020/06/30/home-is-not-a-safe-place-for-everyone-domestic-abuse-between-current-partners-increased-during-lockdown/ [accessed 2021-02-09]

Edited by T Derrick, G Eysenbach; submitted 06.08.20; peer-reviewed by A Azzam, N Mohammad Gholi Mezerji, M K.; accepted 28.01.21; published 17.02.21.

Please cite as:
Grimes HL, Uppal R
Remote Examination and Screening for Domestic Abuse. Comment on "Online Antenatal Care During the COVID-19 Pandemic: Opportunities and Challenges"
J Med Internet Res 2021;23(2):e23295
URL: https://www.jmir.org/2021/2/e23295
doi:10.2196/23295
PMID:33595438

Letter to the Editor

Defining Telemedicine and Engaging Future Medical Practitioners. Comment on “Telemedicine in Germany During the COVID-19 Pandemic: Multi-Professional National Survey”

Neha Sadik[1]; Reem Salman[1]

Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom

Corresponding Author:
Neha Sadik
Barts and the London School of Medicine and Dentistry
Queen Mary University of London
Garrod Building, Turner St
Whitechapel
London, E1 2AD
United Kingdom
Phone: 44 7444321677
Email: ha17302@qmul.ac.uk

Related Article:

Comment on: https://www.jmir.org/2020/8/e19745/

(J Med Internet Res 2021;23(2):e23363) doi:10.2196/23363

KEYWORDS

telemedicine; coronavirus; COVID-19; medical education

We read with interest the recently published paper by Peine et al [1] on the views of health professionals on telemedicine. By surveying health professionals from across a variety of clinical settings and demographics, interesting results were obtained illustrating high levels of acceptance of telemedicine among health professionals in Germany. Acquiring an accurate view of telemedicine has never been more important than during the COVID-19 pandemic, where the usage of telemedicine to communicate safely with patients has increased dramatically.

The authors have made use of the word telemedicine throughout the paper, without specifically defining its meaning in this context. In fact, a 2007 study found up to 104 different peer-reviewed definitions of telemedicine [2] or “eHealth” across many different studies. Thus, we ask the authors of the study to clarify what they meant by telemedicine and if this term was defined for all study participants. We support a definition that encompasses the integration of technology mediums into the delivery of more accessible, improved, and cost-effective health care. The use of telemedicine extends beyond its role as a communication medium; it can be a resource for both health care professionals and patients, an instrument for diagnosis and screening, or a device to globalize medicine.

We note that a strength demonstrated in the survey was the large number of participants and the variety of different health professionals included, making it highly representative of the overall German health professional body. However, now that telemedicine has been incorporated into our day-to-day practice of medicine, it is extremely important to explore specific challenges of telemedicine to individual specialties. For example, recent research has shown that it may be difficult to form a trusting patient-doctor relationship using technology, which may be especially impactful in psychiatry and mental health care [3]. For this reason, the results of this study may reveal additional insight if responses were stratified by specialty. By identifying particular problems, solutions pertaining to each specialty may be discovered. For example, virtual reality (VR) technology that replicates a doctor’s office may help to create an environment where a lasting doctor-patient relationship can be established.

In addition, although a variety of health professionals were surveyed, we feel the views of medical students and clinical tutors are clearly missing from this study. As the first generation to fully experience the intersection of telemedicine and medical education, the opinions of this group could shape the future of medicine and better prepare upcoming doctors for practice in the post–COVID-19 era. By surveying medical students, a demographic that is mostly proficient in technology, we can find more innovative usages of telemedicine for maximal benefit to both doctors and patients.

From this, we conclude that despite the numerous benefits of telemedicine in clinical practice, there is a multitude of challenges to overcome. Major investment into technological infrastructure and training seems to be an unavoidable expense for the advancement of telemedicine. As telehealth becomes an integral part of the new global health care system, we must strive to research and actively work toward its incorporation in health services for efficient patient care at all levels.

Editorial Notice

The corresponding author of "Telemedicine in Germany During the COVID-19 Pandemic: Multi-Professional National Survey" declined our invitation to reply to this commentary.

Conflicts of Interest

None declared.

References

1. Peine A, Paffenholz P, Martin L, Dohmen S, Marx G, Loosen SH. Telemedicine in Germany During the COVID-19 Pandemic: Multi-Professional National Survey. J Med Internet Res 2020 Aug 05;22(8):e19745 [FREE Full text] [doi: 10.2196/19745] [Medline: 32568724]
2. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? A collection of 104 peer-reviewed perspectives and theoretical underpinnings. Telemed J E Health 2007 Oct;13(5):573-590. [doi: 10.1089/tmj.2006.0073] [Medline: 17999619]
3. Hoffmann M, Wensing M, Peters-Klimm F, Szecsenyi J, Hartmann M, Friederich HC, et al. Perspectives of Psychotherapists and Psychiatrists on Mental Health Care Integration Within Primary Care Via Video Consultations: Qualitative Preimplementation Study. J Med Internet Res 2020 Jun 18;22(6):e17569 [FREE Full text] [doi: 10.2196/17569] [Medline: 32554369]

Abbreviations

VR: virtual reality

Edited by T Derrick, G Eysenbach; submitted 10.08.20; peer-reviewed by A Peine; comments to author 19.11.20; revised version received 28.11.20; accepted 28.01.21; published 17.02.21.

Please cite as:
Sadik N, Salman R
Defining Telemedicine and Engaging Future Medical Practitioners. Comment on "Telemedicine in Germany During the COVID-19 Pandemic: Multi-Professional National Survey"
J Med Internet Res 2021;23(2):e23363
URL: https://www.jmir.org/2021/2/e23363
doi:10.2196/23363
PMID:33595454

Letter to the Editor

Quality of Information and Future Directions. Comment on "Influence of Mass and Social Media on Psychobehavioral Responses Among Medical Students During the Downward Trend of COVID-19 in Fujian, China: Cross-Sectional Study"

Smriti Sasikumar[1], BSc; Hafsa Omer Sulaiman[1], BSc; Simran Bedi[2]; Mikhail Nozdrin[3]; Caroline Rundell[4]; Sadia Zaman[3]

[1]St George's University of London, Cranmer Terrace, London, United Kingdom
[2]King's College London, London, United Kingdom
[3]Imperial College, London, United Kingdom
[4]University College London, London, United Kingdom

Corresponding Author:
Smriti Sasikumar, BSc
St George's University of London
Cranmer Terrace
London
United Kingdom
Phone: 44 2086729944
Email: Sasikumar.s@live.sgul.ac.cy

Related Article:

Comment on: https://www.jmir.org/2020/7/e19982/

(J Med Internet Res 2021;23(2):e23168) doi:10.2196/23168

KEYWORDS

COVID-19; social media; mass media; information quality

We read with great interest the recent article by Lin et al [1], which investigated the impact of mass and social media on psychobehavioral responses among medical students in China. We thank the authors for investigating this important issue in such a multifaceted manner. As medical students from four different medical schools in London, we questioned whether certain correlations discussed in the study could be examined further to elucidate the results and allow for some extrapolation. The results from this study are quite intriguing, and we believe this deeper analysis could spark other institutions around the world to examine the powerful effect of media on behaviors during a pandemic.

Since the study is based on Chinese media, we were unfamiliar with the quality of disseminated information related to COVID-19, particularly on social media. According to Ren et al [2], while Western media tends to center around controversy and individual experts, Chinese media sources focus on "mainstream positions" put forth by health agencies [3,4]. We suggest that an overview of how media sources are presented in China would be extremely useful in contextualizing the results of the study for non-Chinese audiences.

Furthermore, while the article's findings suggest that exposure to mass media and social media could increase positive attitudes and reduce emotional response among university students, we believe that the credibility of the information on these platforms greatly contributes to the results. Misinformation and "fake news," particularly on the internet, have become rampant [5]. The article presumes that the results are due to medical students being more discerning consumers of health information. While this may be possible, we suggest that analyzing the quality of the information consumed by the students, that is, whether the information was credible or not, would allow stronger correlations to be drawn as to why students had positive behavioral responses and lower trait anxiety.

This analysis on the quality of mass and social media could be done in a number of ways. For example, examining the quality of data could involve recruiting some participants from the current study to respond to health information of differing qualities. The information could be put into two categories: credible information and misinformation about COVID-19. Repeating the STAI-6 (Spielberger State-Trait Anxiety Inventory, 6 items) and the psychobehavioral response

questionnaire on this controlled set of stimuli might yield more data on why mass and social media exposure reduced trait anxiety and changed behavior in the study population.

Lastly, we suggest that the data be further analyzed to examine whether the psychobehavioral and anxiety responses varied by age in the students. This suggestion stems from the observed trend that different age groups engage with different mass or social media platforms. Perhaps the quality and the type of mass or social media may have an effect on the observed results.

Implementing some of these suggestions could use the previously gathered data and provide a uniquely detailed viewpoint that could expand the results of the study. These additions and future steps could inspire cross-institutional collaboration, allowing us to better understand how media across the world can affect the behavior and anxiety levels of key populations during a pandemic.

Editorial Notice

The corresponding author of "Influence of Mass and Social Media on Psychobehavioral Responses Among Medical Students During the Downward Trend of COVID-19 in Fujian, China: Cross-Sectional Study" did not respond to our invitation to reply to this commentary.

Conflicts of Interest

None declared.

References

1. Lin Y, Hu Z, Alias H, Wong LP. Influence of Mass and Social Media on Psychobehavioral Responses Among Medical Students During the Downward Trend of COVID-19 in Fujian, China: Cross-Sectional Study. J Med Internet Res 2020 Jul 20;22(7):e19982 [FREE Full text] [doi: 10.2196/19982] [Medline: 32584779]
2. Ren J, Peters HP, Allgaier J, Lo YY. Similar challenges but different responses: Media coverage of measles vaccination in the UK and China. Public Underst Sci 2014 May;23(4):366-375 [FREE Full text] [doi: 10.1177/0963662512445012] [Medline: 23825271]
3. DeGennaro T. 10 Most Popular Social Media Sites in China 2019. Dragon Social. 2020 Jun 30. URL: https://www.dragonsocial.net/blog/social-media-in-china/ [accessed 2021-02-09]
4. Social Media Fact Sheet 2019. Pew Research Center. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2021-02-09]
5. Sommariva S, Vamos C, Mantzarlis A, Đào L, Martinez Tyson D. Spreading the (Fake) News: Exploring Health Messages on Social Media and the Implications for Health Professionals Using a Case Study. American Journal of Health Education 2018 Jun 07;49(4):246-255. [doi: 10.1080/19325037.2018.1473178]

Abbreviations

STAI-6: Spielberger State-Trait Anxiety Inventory, 6 items

Edited by G Eysenbach, T Derrick; submitted 03.08.20; peer-reviewed by J Li, S Temiz, M K.; accepted 28.01.21; published 18.02.21.

Please cite as:
Sasikumar S, Sulaiman HO, Bedi S, Nozdrin M, Rundell C, Zaman S
Quality of Information and Future Directions. Comment on "Influence of Mass and Social Media on Psychobehavioral Responses Among Medical Students During the Downward Trend of COVID-19 in Fujian, China: Cross-Sectional Study"
J Med Internet Res 2021;23(2):e23168
URL: https://www.jmir.org/2021/2/e23168
doi:10.2196/23168
PMID:33599623

Corrigenda and Addenda

Correction: Estimated Sleep Duration Before and During the COVID-19 Pandemic in Major Metropolitan Areas on Different Continents: Observational Study of Smartphone App Data

Rebecca Robbins[1,2], PhD; Mahmoud Affouf[3], PhD; Matthew D Weaver[1,2], PhD; Mark É Czeisler[4,5], AB; Laura K Barger[1,2], PhD; Stuart F Quan[1,2], MD; Charles A Czeisler[1,2], MD, PhD, FRCP

[1]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA, United States

[2]Division of Sleep Medicine, Harvard Medical School, Boston, MA, United States

[3]Department of Mathematics, Kean University, Union, NJ, United States

[4]School of Psychological Sciences, Turner Institute Brain and Mental Health, Monash University, Victoria, Australia

[5]Institute for Breathing and Sleep, Austin Health, Melbourne, Australia

Corresponding Author:
Rebecca Robbins, PhD
Division of Sleep and Circadian Disorders
Departments of Medicine and Neurology
Brigham and Women's Hospital
221 Longwood Avenue
Boston, MA, 02115
United States
Phone: 1 2039792338
Email: rrobbins4@bwh.harvard.edu

Related Article:

Correction of: https://www.jmir.org/2021/2/e20546/

(J Med Internet Res 2021;23(2):e28057) doi:10.2196/28057

In "Estimated Sleep Duration Before and During the COVID-19 Pandemic in Major Metropolitan Areas on Different Continents: Observational Study of Smartphone App Data" (J Med Internet Res 2021;23(2):e20546) the authors noted one error.

In the originally published paper, the name of one author cited in Reference 25 (Czeisler MÉ) was incomplete. The list of authors cited in Reference 25 originally appeared as follows:

> *Czeisler M, Howard ME, Robbins R, Barger LK, Facer-Childs ER, Rajaratnam SM, et al.*

In the corrected version of the paper, the list of authors appears as follows:

> *Czeisler MÉ, Howard ME, Robbins R, Barger LK, Facer-Childs ER, Rajaratnam SM, et al.*

The correction will appear in the online version of the paper on the JMIR Publications website on February 22, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 18.02.21; this is a non–peer-reviewed article;accepted 18.02.21; published 22.02.21.

Please cite as:
Robbins R, Affouf M, Weaver MD, Czeisler MÉ, Barger LK, Quan SF, Czeisler CA
Correction: Estimated Sleep Duration Before and During the COVID-19 Pandemic in Major Metropolitan Areas on Different Continents: Observational Study of Smartphone App Data
J Med Internet Res 2021;23(2):e28057
URL: https://www.jmir.org/2021/2/e28057
doi:10.2196/28057
PMID:33617457

Corrigenda and Addenda

Correction: Adequacy of Web-Based Activities as a Substitute for In-Person Activities for Older Persons During the COVID-19 Pandemic: Survey Study

Jiska Cohen-Mansfield[1,2,3], PhD; Aline Muff[2], PhD; Guy Meschiany[2], MA; Shahar Lev-Ari[1], PhD

[1]Department of Health Promotion, School of Public Health, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

[2]Minerva Center for the Interdisciplinary Study of End of Life, Tel Aviv University, Tel Aviv, Israel

[3]Igor Orenstein Chair for the Study of Geriatrics, Tel Aviv University, Tel Aviv, Israel

Corresponding Author:
Jiska Cohen-Mansfield, PhD
Department of Health Promotion
School of Public Health, Sackler Faculty of Medicine
Tel Aviv University
POB 39040
Tel Aviv
Israel
Phone: 972 03 6407576
Email: jiska@tauex.tau.ac.il

Related Article:

Correction of: https://www.jmir.org/2021/1/e25848/

(J Med Internet Res 2021;23(2):e27687) doi:10.2196/27687

In "Adequacy of Web-Based Activities as a Substitute for In-Person Activities for Older Persons During the COVID-19 Pandemic: Survey Study" (J Med Internet Res 2021;23(1):e25848) the authors noted one error.

In the originally published article, the first sentence of the final paragraph of the section "Sustainability: Competing Activities and Willingness to Pay for Activities" read as follows:

> *Out of the 56 nonparticipants, 20 (36%) indicated willingness to pay for Zoom activities; however, 12 (26%) responded in the negative.*

This was incorrect, and has been changed to:

> *Out of the 56 nonparticipants, 28 (50%) indicated willingness to pay for Zoom activities; however, 20 (36%) responded in the negative.*

The correction will appear in the online version of the paper on the JMIR Publications website on February 16, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 02.02.21; this is a non–peer-reviewed article;accepted 02.02.21; published 16.02.21.

Please cite as:
Cohen-Mansfield J, Muff A, Meschiany G, Lev-Ari S
Correction: Adequacy of Web-Based Activities as a Substitute for In-Person Activities for Older Persons During the COVID-19 Pandemic: Survey Study
J Med Internet Res 2021;23(2):e27687
URL: http://www.jmir.org/2021/2/e27687/
doi:10.2196/27687
PMID:33591925

Viewpoint

Telehealth in the COVID-19 Era: A Balancing Act to Avoid Harm

J Jeffery Reeves[1*], MD; John W Ayers[2*], PhD; Christopher A Longhurst[2*], MD

[1]Department of Surgery, University of California San Diego, La Jolla, CA, United States

[2]Department of Medicine, Division of Biomedical Informatics, University of California San Diego, La Jolla, CA, United States

[*]all authors contributed equally

Corresponding Author:
J Jeffery Reeves, MD
Department of Surgery
University of California San Diego
9300 Campus Point Drive, MC7400
La Jolla, CA, 92037-7400
United States
Phone: 1 505 515 9844
Email: jreeves@ucsd.edu

Abstract

The telehealth revolution in response to COVID-19 has increased essential health care access during an unprecedented public health crisis. However, virtual patient care can also limit the patient-provider relationship, quality of examination, efficiency of health care delivery, and overall quality of care. As we witness the most rapidly adopted medical trend in modern history, clinicians are beginning to comprehend the many possibilities of telehealth, but its limitations also need to be understood. As outcomes are studied and federal regulations reconsidered, it is important to be precise in the virtual patient encounter approach. Herein, we offer some simple guidelines that could assist health care providers and clinic schedulers in determining the appropriateness of a telehealth visit by considering visit types, patient characteristics, and chief complaint or disease states.

(J Med Internet Res 2021;23(2):e24785) doi:10.2196/24785

KEYWORDS

telehealth; patient safety; COVID-19; coronavirus; informatics; safety; harm; risk; access; efficiency; virtual care

Introduction

The telehealth revolution has been heralded for its potential to increase health care access and improve the efficiency of health care delivery. Previously, telehealth had been strategically rolled out to settings wherein the need for a relationship is less pressing (eg, radiology), outcomes can be remotely assessed in principle (eg, dermatology), or access to care is severely limited (eg, rural areas) [1]. COVID-19 prompted a near-universal expansion of telehealth utilization. Many applaud this change [2]. However, virtual patient care can bring unintended consequences that eclipse its benefits, including potentially limiting the patient-provider relationship, quality of examination, the efficiency of health care delivery, and overall quality of care [3-6]. As safety protocols for in-person evaluation have been developed, critical examination raises serious dilemmas about the harms of telehealth that can inform a new agenda for how to best utilize telehealth in the COVID-19 era. As we witness the most rapidly adopted medical trend in modern history, clinicians are beginning to comprehend the possibilities of telehealth, but its limitations also need to be understood.

The Rise of Telehealth in Response to COVID-19

Around the globe, patients are wary of entering health care settings, and providers fear unnecessary exposure for both patients and themselves, thereby creating a familiar dilemma of an unprecedented variety: How to best care for patients while *first doing no harm?*

In response, health systems have rapidly transitioned patient care away from in-person encounters towards telehealth [7]. This transition was facilitated by legislative changes designed to enable the “good faith” provision of health care at a distance, thereby limiting COVID-19 transmission [8]. New reimbursement policies have expanded access to a range of telehealth services offered by medical providers, clinical psychologists, licensed clinical social workers, and other health care workers.

In a matter of weeks, the application of patient-facing technology spread across all outpatient settings, including primary and preventative, medical and surgical specialty, and

mental health care. Equipped with remote monitoring, access to multi-provider video visits, and virtual translators, some specialty clinics adopted a 100% virtual approach. Telehealth found its way into emergency departments, hospital wards, intensive care units, and even interdisciplinary services such as occupational and physical therapy [9]. Telehealth was originally regarded as a form of health care delivery largely reserved for specific, resource-limited settings, but it is now embedded in the daily practice of providers across the spectrum of patient care.

Unintended Consequences of Telehealth

First, the most obvious potential harm of telehealth is an incomplete or inaccurate physical examination. Advances in patients' technological assessments have historically led to the abandonment of physical touch and examination in general—COVID-19 potentiated this issue [10]. Several reports have documented the steps to perform a virtual physical examination that can be applied to a wide range of patient encounters [11]. For example, Tanaka et al [12] describe creative ways of performing a virtual orthopedic examination. However, there are real limitations to a virtual examination, including the potential requirement of assistance from a caregiver, inadequate patient-home environment or space, poor lighting or discoloration with resultant poor visualization, increased time required to perform an assessment, and other technical issues [6,13,14]. In particular, the assessment of a patient's movement and motor function is limited, in part due to the need for a patient-assisted examination that can be challenging to capture on video [6,15,16]. Other aspects of a physical examination simply cannot be done via telehealth, such as auscultation of heart or lung sounds [17]. Thus, the true effectiveness of a telehealth physical examination and its subsequent impact on diagnosis, clinical management, and medical outcomes is yet to be determined in all settings.

In some instances, health care providers have abandoned the physical aspect of the "history and physical." The Centers for Medicare and Medicaid Services reimburse for telehealth annual well visits—an encounter that historically offered providers an opportunity to perform a thorough examination for subtle abnormalities is now often performed purely on the telephone. In other settings, reliance on technology has replaced physical examination. Inpatient cardiology consults are now performed by providers using apps on their devices (eg, iPhones and iPads) that allow real-time electrocardiography with an acceptable neglect of a careful, in-person examination of neck veins and body tissue to assess fluid status.

Beyond the clinical value, many patients feel comforted and reassured after a physical examination. Others perceive a virtual examination as inadequate regardless of its true quality [18]. "We must not forget the importance of a doctor's hands" [19]—a phrase that used to symbolize compassion is now, based on Google search results, associated with COVID-19 safety information and online advice to self-isolate.

Second, the roll-out of telehealth undermines the patient-provider relationship and the essential humanistic qualities of care providers. Although the world has become accustomed to online relationships, a trusting and personal connection between the patient and the health care provider, a vital aspect of health care, can be challenging to build with purely digital interaction. Multiple patient experience studies have consistently reported difficulties in communicating or connecting with providers during telehealth visits [3,6]. A desire for building an improved rapport with their providers is one of the most popular reasons why many patients prefer in-person visits [20]. In particular, the establishment of primary care or the initial consultation, similar to a first date, may not be welcomed by all and could negatively affect patient experience or the development of a sense of "my doctor." A survey of primary care patients during the pandemic found that telehealth was considered the most appropriate in the presence of a pre-existing clinical relationship [21]. An analysis of 620 patient satisfaction survey outcomes from clinic encounters found that new visit type was associated with lower patient satisfaction (parameter estimate -0.75; 95% CI -1.00 to -0.049) [22]. Consider the impact on the patient-centered care model, in which empathy, two-way communication, and direct eye-to-eye contact are considered crucial elements to improving health literacy and engaging patients in their own disease management [23]. Depending on the situation, telehealth may add one more barrier to understanding complex medical disease and patient compliance.

Third, the expansion of virtual services to rates beyond what was previously expected has resulted in inefficiencies in health care delivery. A mass of health care organizations rapidly transitioned and onboarded providers to telehealth encounters in a matter of weeks during the early-stage of the COVID-19 pandemic [24]. Although telehealth can provide thorough, safe, and effective medical care remotely, there is a learning curve associated with its implementation, and its success requires logistics that are frequently overlooked. For instance, clinics have not been appropriately restructured to support telehealth visits, resulting in under-utilization of talent. For example, medical assistants now play a role similar to that of front-desk personnel, thereby considerably limiting their participation in patient care. As such, the existing infrastructure, education, and administrative support surrounding telehealth must be tailored and broadened.

Safety of Health Care Facilities

The rapid expansion of telehealth was intended to protect uninfected patients from the risks of COVID-19 during periods of uncertainty. Although disease transmission is known to occur in medical facilities, growing evidence suggests that health care workers are more likely to be exposed to the virus while performing extra-occupational activities [25,26]. More experience in handling COVID-19 cases, coupled with an increasing knowledge base of its characteristics and modes of transmission, has led to the establishment of evidence-based protocols proven to facilitate the safe delivery of traditional brick-and-mortar patient care [27,28]. Retrospective studies have shown that transmission of the virus among health care workers is considerably reduced when these protocols are followed [29]. Proper hand hygiene, facial coverings, use of appropriate personal protective equipment, and implementation

of social distancing practices allow the safe return of routine in-person services [30]. Additionally, the redesign of clinic spaces and the use of electronic registration with automated messaging to avoid crowded waiting areas can further enhance the safety of in-person patient encounters. As such, providers now can be more precise in which patients are seen physically rather than virtually.

Updating Telehealth for the COVID-19 Era

For certain encounters, it makes intuitive sense to conduct telehealth visits remotely, and this may ultimately become the standard of care in the wake of the COVID-19 pandemic. For example, during a medication review for patients with chronic medical conditions, a physician can ask the patient to point the camera to the labels of the pills they are actually taking. On the other hand, some chief complaints require a physical encounter—a newly recognized palpable mass, for instance. The problem lies in the equivocal encounters: the patient who keeps physicians up at night, making them wonder if they should have recommended a trip to the dreaded emergency department, or the well-child, who is up to date on immunizations but has not had their weight or development assessed. As clinics reopen, finding the optimal balance of in-person and telehealth patient encounters is one of the most critical questions facing health care delivery. Appropriateness of telehealth visits likely varies according to the type of service being offered, medical specialty, health system, and geographic location.

We offer some simple guidelines to consider in determining the appropriateness of a telehealth visit (Table 1); these guidelines may be utilized by health care providers and, importantly, clinic schedulers. First, the visit type should be considered. New patient visits, even those that perhaps are safe to perform virtually, may benefit from face-to-face encounters to familiarize patients with their care and build a trusting relationship. Follow-up visits or new patients unlikely to need a physical examination may be appropriately managed via telehealth. Second, many patient characteristics are important to be considered, including health literacy, the structure of social support systems, hearing ability or visual acuity, and simple preference. A simple assessment of these 2 factors can likely be made by clinic schedulers with no further context of a patient's medical history. Thus, appropriate delineation of telehealth utilization in the outpatient setting can be made primarily without the consultation of a provider. However, for borderline patients, a provider can consider the chief complaint and chronicity of disease state as key determinants of whether in-person care with a highly reliable physical examination is necessary. When a physical examination may change either the diagnosis or management strategy, the patient should be seen in-person. This distinction can often be determined based on the chief complaint [31]. Finally, the technological capabilities of the patient must not be overlooked or taken for granted.

In the future, precision in the approach and delivery of the telehealth patient encounter is essential. The telehealth revolution has been impressive and has made substantial contributions to the global management of this pandemic, but significant work remains to ensure the best health care delivery for our patients. We have an opportunity to simultaneously embrace the necessity and benefits of telehealth while supporting strategies that optimize patient outcomes and humanism in medicine. Universally, health care workers have labored under extraordinary circumstances with incredible and heart-warming fortitude. Front-line providers have reminded us of the virtue of health care. There is no certainty about what the postpandemic world will look like. However, there will always be patients in need, and it is our responsibility to ensure they receive the high-quality physical and personal care they deserve.

Table 1. Characteristics to consider for determining the appropriateness of telehealth.

Characteristic	Appropriate for telehealth	Potentially inappropriate for telehealth
Visit type	Follow-up visit for known/diagnosed disease state or patient condition	New patient (establishment of primary care or new consultation)
	Follow-up postprocedure visit with no patient complaints	Annual physical examination or well-child check
	Recurring medication or chronic medical condition review	Acute visit prompted by an acute change in patient condition
	Initial or follow-up visit for mental health conditions	Initial psychiatry visits or annual follow-up for controlled substances
Patient characteristics	Existing, trusting, personal connection with the provider	Distrusting of health care professionals
	High health literacy	Low health literacy
	Robust social support system	Low social support system
	Anxious in health care settings	Poor vision or hearing
	Lives remotely or has inadequate transportation	Prefers in-person
	Prefers telehealth	
Chief complaint or disease state characteristics	Physical examination unlikely to be diagnostic	Physical examination may aid in diagnosis or prognosis
	N/A[a]	Examination findings may influence initial workup and/or management
	Focused physical examination can be performed virtually (ie, visual examination)	Focused examination cannot be performed virtually (ie, palpation of mass)
	Chief complaint with standardized initial workup and management	Chief complaints that often result in referral to acute care settings
Other considerations	English as a second language (interpreter required)	Patient has a poor internet connection
	Patient or close family member technologically savvy	Patient lacks technological capabilities to join video visit/telehealth encounter
	Patient with multiple family members at home who can join telehealth visit	N/A

[a]N/A: not applicable.

Authors' Contributions

CAL conceived the work. JJR, JWA, and CAL contributed to the design and content of the manuscript. JJR wrote the initial manuscript. JWA and CAL critically reviewed and substantially revised the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

JWA owns equity positions in Directing Medicine, Health Watcher, and Good Analytics, which are companies that advise on the use of digital data for public health surveillance. All other authors have no conflicts to declare.

References

1. Moore MA, Coffman M, Jetty A, Klink K, Petterson S, Bazemore A. Family physicians report considerable interest in, but limited use of, telehealth services. J Am Board Fam Med 2017 May 08;30(3):320-330. [doi: 10.3122/jabfm.2017.03.160201]
2. Mehrotra A, Ray K, Brockmeyer D, Barnett M, Bender J. Rapidly converting to “virtual practices”: outpatient care in the era of Covid-19. NEJM Catal Innov Care Deliv 2020 Apr;1(2). [doi: 10.1056/CAT.20.0091]
3. Mair F, Whitten P. Systematic review of studies of patient satisfaction with telemedicine. BMJ 2000 Jun 03;320(7248):1517-1520 [FREE Full text] [doi: 10.1136/bmj.320.7248.1517] [Medline: 10834899]
4. Wachter R. The digital doctor: hope, hype, and harm at the dawn of medicine’s computer age. Crit Care Nurse 2016;36(4):84-85. [doi: 10.4037/ccn2016880]
5. Dorsey ER, Topol EJ. State of telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]

6. Saliba-Gustafsson EA, Miller-Kuhlmann R, Kling SMR, Garvert DW, Brown-Johnson CG, Lestoquoy AS, et al. Rapid implementation of video visits in neurology during COVID-19: mixed methods evaluation. J Med Internet Res 2020 Dec 09;22(12):e24328 [FREE Full text] [doi: 10.2196/24328] [Medline: 33245699]
7. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]
8. Shachar C, Engel J, Elwyn G. Implications for telehealth in a postpandemic future: regulatory and privacy issues. JAMA 2020 Jun 16;323(23):2375-2376. [doi: 10.1001/jama.2020.7943] [Medline: 32421170]
9. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82 [FREE Full text] [doi: 10.1056/nejmp2005835]
10. Horton R. Offline: touch—the first language. The Lancet 2019 Oct;394(10206):1310. [doi: 10.1016/s0140-6736(19)32280-9]
11. Benziger CP, Huffman MD, Sweis RN, Stone NJ. The telehealth ten: a guide for a patient-assisted virtual physical examination. Am J Med 2021 Jan;134(1):48-51 [FREE Full text] [doi: 10.1016/j.amjmed.2020.06.015] [Medline: 32687813]
12. Tanaka MJ, Oh LS, Martin SD, Berkson EM. Telemedicine in the era of COVID-19: the virtual orthopaedic examination. J Bone Joint Surg Am 2020 Jun 17;102(12):e57 [FREE Full text] [doi: 10.2106/JBJS.20.00609] [Medline: 32341311]
13. Mammen JR, Elson MJ, Java JJ, Beck CA, Beran DB, Biglan KM, et al. Patient and physician perceptions of virtual visits for Parkinson's disease: a qualitative study. Telemed J E Health 2018 Apr;24(4):255-267. [doi: 10.1089/tmj.2017.0119] [Medline: 28787250]
14. Kemp MT, Liesman DR, Williams AM, Brown CS, Iancu AM, Wakam GK, et al. Surgery provider perceptions on telehealth visits during the COVID-19 pandemic: room for improvement. J Surg Res 2020 Nov 13;260:300-306 [FREE Full text] [doi: 10.1016/j.jss.2020.11.034] [Medline: 33360755]
15. Hoenig H, Tate L, Dumbleton S, Montgomery C, Morgan M, Landerman LR, et al. A quality assurance study on the accuracy of measuring physical function under current conditions for use of clinical video telehealth. Arch Phys Med Rehabil 2013 May;94(5):998-1002. [doi: 10.1016/j.apmr.2013.01.009] [Medline: 23337425]
16. Yager PH, Clark ME, Dapul HR, Murphy S, Zheng H, Noviski N. Reliability of circulatory and neurologic examination by telemedicine in a pediatric intensive care unit. J Pediatr 2014 Nov;165(5):962-6.e1. [doi: 10.1016/j.jpeds.2014.07.002] [Medline: 25112695]
17. Chowdhury D, Hope KD, Arthur LC, Weinberger SM, Ronai C, Johnson JN, et al. Telehealth for pediatric cardiology practitioners in the time of COVID-19. Pediatr Cardiol 2020 Aug;41(6):1081-1091 [FREE Full text] [doi: 10.1007/s00246-020-02411-1] [Medline: 32656626]
18. Powell RE, Henstenburg JM, Cooper G, Hollander JE, Rising KL. Patient perceptions of telehealth primary care video visits. Ann Fam Med 2017 May;15(3):225-229 [FREE Full text] [doi: 10.1370/afm.2095] [Medline: 28483887]
19. Verghese A. Culture shock — patient as icon, icon as patient. N Engl J Med 2008 Dec 25;359(26):2748-2751. [doi: 10.1056/nejmp0807461]
20. Lacritz L, Carlew A, Livingstone J, Bailey K, Parker A, Diaz A. Patient satisfaction with telephone neuropsychological assessment. Arch Clin Neuropsychol 2020 Nov 19;35(8):1240-1248 [FREE Full text] [doi: 10.1093/arclin/acaa097] [Medline: 33124648]
21. Imlach F, McKinlay E, Middleton L, Kennedy J, Pledger M, Russell L, et al. Telehealth consultations in general practice during a pandemic lockdown: survey and interviews on patient experiences and preferences. BMC Fam Pract 2020 Dec 13;21(1):269 [FREE Full text] [doi: 10.1186/s12875-020-01336-1] [Medline: 33308161]
22. Ramaswamy A, Yu M, Drangsholt S, Ng E, Culligan PJ, Schlegel PN, et al. Patient satisfaction with telemedicine during the COVID-19 pandemic: retrospective cohort study. J Med Internet Res 2020 Sep 09;22(9):e20786 [FREE Full text] [doi: 10.2196/20786] [Medline: 32810841]
23. What Is Patient-Centered Care? NEJM Catal. NEJM Catalyst 2017;3(1):1-1.
24. Patel PD, Cobb J, Wright D, Turer R, Jordan T, Humphrey A, et al. Rapid development of telehealth capabilities within pediatric patient portal infrastructure for COVID-19 care: barriers, solutions, results. J Am Med Inform Assoc 2020 Jul 01;27(7):1116-1120 [FREE Full text] [doi: 10.1093/jamia/ocaa065] [Medline: 32302395]
25. Lan F, Wei C, Hsu Y, Christiani DC, Kales SN. Work-related COVID-19 transmission in six Asian countries/areas: A follow-up study. PLoS One 2020;15(5):e0233588 [FREE Full text] [doi: 10.1371/journal.pone.0233588] [Medline: 32428031]
26. Lentz RJ, Colt H, Chen H, Cordovilla R, Popevic S, Tahura S, et al. Assessing coronavirus disease 2019 (COVID-19) transmission to healthcare personnel: The global ACT-HCP case-control study. Infect Control Hosp Epidemiol 2020 Sep 09:1-7 [FREE Full text] [doi: 10.1017/ice.2020.455] [Medline: 32900402]
27. Centers FDC. Following Infection Prevention and Control Recommendations in Healthcare Settings During the COVID-19 Pandemic. Healthcare Facilities: Managing Operations During the COVID-19 Pandemic. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/guidance-hcf.html [accessed 2020-10-01]
28. Gan WH, Lim JW, Koh D. Preventing intra-hospital infection and transmission of coronavirus disease 2019 in health-care workers. Saf Health Work 2020 Jun;11(2):241-243 [FREE Full text] [doi: 10.1016/j.shaw.2020.03.001] [Medline: 32292622]
29. Coronavirus disease 2019 (COVID-19): situation report, 82. World Health Organization. 2020. URL: https://apps.who.int/iris/handle/10665/331780 [accessed 2021-01-21]

30. Interim Infection Prevention and Control Recommendations for Healthcare Personnel During the Coronavirus Disease 2019 (COVID-19) Pandemic. Centers for Disease Control and Prevention - COVID-19. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/infection-control-recommendations.html [accessed 2020-09-03]
31. Gadzinski AJ, Ellimoottil C. Telehealth in urology after the COVID-19 pandemic. Nat Rev Urol 2020 Jul;17(7):363-364 [FREE Full text] [doi: 10.1038/s41585-020-0336-6] [Medline: 32405032]

Edited by G Eysenbach; submitted 05.10.20; peer-reviewed by J Hefner, D Bychkov; comments to author 18.11.20; revised version received 20.12.20; accepted 16.01.21; published 01.02.21.

Please cite as:
Reeves JJ, Ayers JW, Longhurst CA
Telehealth in the COVID-19 Era: A Balancing Act to Avoid Harm
J Med Internet Res 2021;23(2):e24785
URL: https://www.jmir.org/2021/2/e24785
doi:10.2196/24785
PMID:33477104

Original Paper

The Use of Digital Platforms for Adults' and Adolescents' Physical Activity During the COVID-19 Pandemic (Our Life at Home): Survey Study

Kate Parker[1], PhD; Riaz Uddin[1], PhD; Nicola D Ridgers[1], PhD; Helen Brown[1], PhD; Jenny Veitch[1], PhD; Jo Salmon[1], PhD; Anna Timperio[1], PhD; Shannon Sahlqvist[1], PhD; Samuel Cassar[1], MA; Kim Toffoletti[2], PhD; Ralph Maddison[1], PhD; Lauren Arundell[1], PhD

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

[2]School of Humanities and Social Sciences, Faculty of Arts and Education, Deakin University, Burwood, Australia

Corresponding Author:
Kate Parker, PhD
Institute for Physical Activity and Nutrition
Deakin University
75 Pigdons Rd
Geelong
Australia
Phone: 61 92468094
Email: k.parker@deakin.edu.au

Abstract

Background: Government responses to managing the COVID-19 pandemic may have impacted the way individuals were able to engage in physical activity. Digital platforms are a promising way to support physical activity levels and may have provided an alternative for people to maintain their activity while at home.

Objective: This study aimed to examine associations between the use of digital platforms and adherence to the physical activity guidelines among Australian adults and adolescents during the COVID-19 stay-at-home restrictions in April and May 2020.

Methods: A national online survey was distributed in May 2020. Participants included 1188 adults (mean age 37.4 years, SD 15.1; 980/1188, 82.5% female) and 963 adolescents (mean age 16.2 years, SD 1.2; 685/963, 71.1% female). Participants reported demographic characteristics, use of digital platforms for physical activity over the previous month, and adherence to moderate- to vigorous-intensity physical activity (MVPA) and muscle-strengthening exercise (MSE) guidelines. Multilevel logistic regression models examined differences in guideline adherence between those who used digital platforms (ie, users) to support their physical activity compared to those who did not (ie, nonusers).

Results: Digital platforms include streaming services for exercise (eg, YouTube, Instagram, and Facebook); subscriber fitness programs, via an app or online (eg, Centr and MyFitnessPal); facilitated online live or recorded classes, via platforms such as Zoom (eg, dance, sport training, and fitness class); sport- or activity-specific apps designed by sporting organizations for participants to keep up their skills (eg, TeamBuildr); active electronic games (eg, Xbox Kinect); and/or online or digital training or racing platforms (eg, Zwift, FullGaz, and Rouvy). Overall, 39.5% (469/1188) of adults and 26.5% (255/963) of adolescents reported using digital platforms for physical activity. Among adults, MVPA (odds ratio [OR] 2.0, 95% CI 1.5-2.7), MSE (OR 3.3, 95% CI 2.5-4.5), and combined (OR 2.7, 95% CI 2.0-3.8) guideline adherence were higher among digital platform users relative to nonusers. Adolescents' MVPA (OR 2.4, 95% CI 1.3-4.3), MSE (OR 3.1, 95% CI 2.1-4.4), and combined (OR 4.3, 95% CI 2.1-9.0) guideline adherence were also higher among users of digital platforms relative to nonusers.

Conclusions: Digital platform users were more likely than nonusers to meet MVPA and MSE guidelines during the COVID-19 stay-at-home restrictions in April and May 2020. Digital platforms may play a critical role in helping to support physical activity engagement when access to facilities or opportunities for physical activity outside the home are restricted.

(J Med Internet Res 2021;23(2):e23389) doi:10.2196/23389

KEYWORDS

digital health; moderate- to vigorous-intensity physical activity; muscle-strengthening exercise; online platforms; COVID-19

Introduction

Physical activity plays an important role in the prevention and treatment of noncommunicable diseases, which account for 70% of deaths worldwide [1]. A recent study of 168 countries estimated that 3.9 million (15%) premature deaths could be averted annually if more people engaged in recommended levels of physical activity [2]. The World Health Organization recommends that, each week, adults should accumulate at least 150 minutes of moderate-intensity physical activity, 75 minutes of vigorous-intensity physical activity, or an equivalent combination of moderate- to vigorous-intensity physical activity (MVPA); in addition, youth aged 5 to 17 years should accumulate at least 60 minutes of MVPA daily [3]. Adults are also advised to perform muscle-strengthening exercises (MSEs) at least twice per week, and adolescents are advised to perform them at least three times per week [3]. The guidelines are consistent with those in Australia [4,5]. Globally, 19% of adolescents [6] and 73% of adults achieve the MVPA guidelines according to the most recent estimates [7]. National survey data from 2017-2018 show that fewer Australians engaged in the recommended physical activity levels than the global average, with 55% of adults (18-64 years) and 10% of adolescents (15-17 years) achieving the recommended MVPA [8], and 1 in 6 adults (15%) and adolescents (16%) adhering to the MSE guidelines [8]. These prevalence data have been observed in the general population under free-living conditions. However, conditions have changed considerably as a result of government responses to COVID-19, which led to unprecedented and widespread social distancing measures to control its spread [9].

In Australia, for example, the federal government announced strict *stay-at-home* orders in late March 2020. Although the length of these restrictions varied by state and territory, opportunities to perform some physical activities outside the home, such as at gyms and sport clubs, were impacted nationwide [9,10]. Google Trends data showed that online queries of how to perform physical activity and exercise peaked in Australia during the first 2 weeks that the stay-at-home restrictions were imposed [11]. While this shows that people were investigating ways to keep active during this period, it does not provide information on the sort of support or the types of programs people may have used to be active during this time. Over the same period, there was an exponential increase in the use of the internet and associated digital platforms, such as websites and smartphone apps, as they became essential for education, work, and social interactions [12]. Digital platforms have previously shown promise for increasing physical activity among individuals of all ages in intervention studies [13-15]. However, use among Australians prior to the pandemic was low. Surveys of Australian adults (≥15 years of age) in 2018 showed that just 18.7% of adults used apps for tracking activity or training, and engagement with websites or online tools (7.1%) and online videos for sport (2.5%) was low [16].

Understanding the types of digital platforms used during the unprecedented pandemic situation could provide insight into their potential role for supporting individuals in meeting MVPA, MSE, and combined physical activity recommendations when individuals are unable to access traditional physical activity settings and facilities. This study aimed to explore the use of digital platforms for physical activity in Australia during April and May 2020 and to examine associations between the use of digital platforms and adherence to physical activity guidelines among adults and adolescents.

Methods

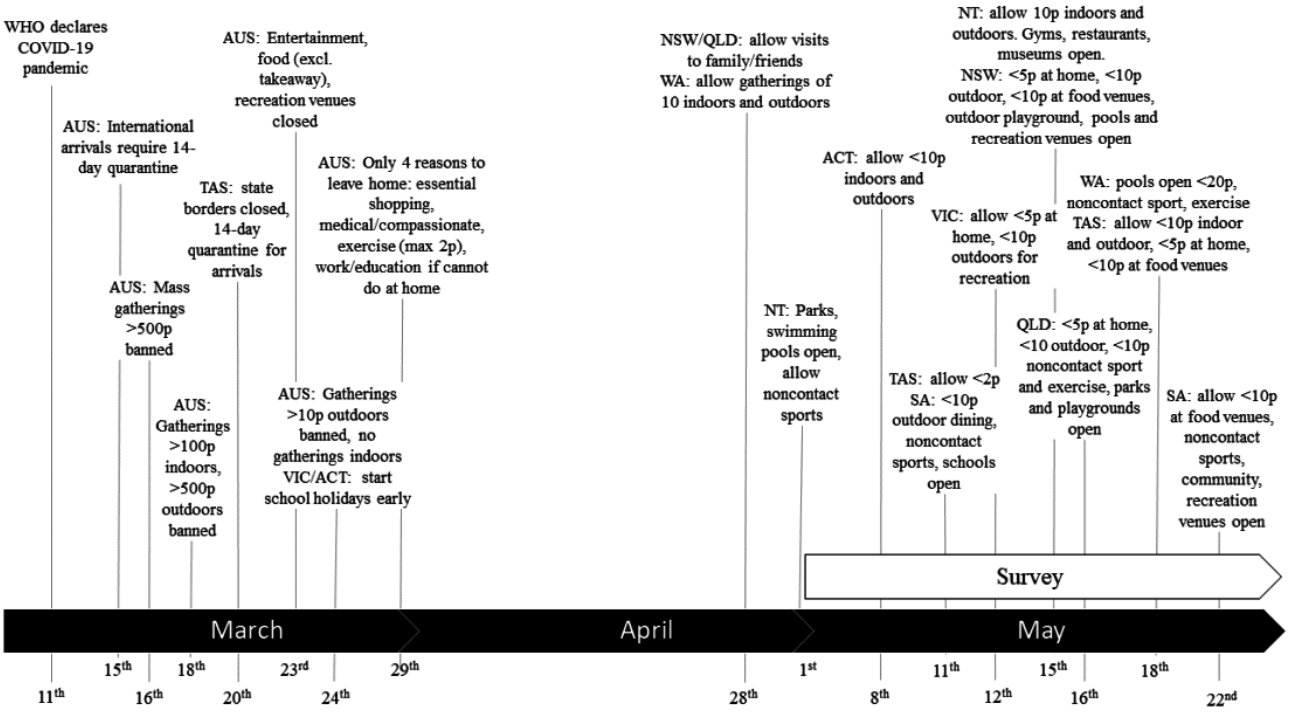
Overview

Data were drawn from the baseline sample of the Our Life at Home study (OL@H), collected May 4-31, 2020. OL@H is a longitudinal study designed to investigate the impact of the Australian Governments' response to managing the COVID-19 pandemic on movement behaviors and on the health and well-being of Australians aged 13 to 75 years over a 2-year period. OL@H received ethical approval from the Deakin University Human Ethics Advisory Group-Health (HEAG-H 59_2020).

Context

From March 29, 2020, strict stay-at-home orders were imposed by the Australian Government. As a result, all organized and social sports were suspended, and gyms and recreation facilities temporarily closed. People were allowed to leave their home to exercise as long as they maintained a 1.5-meter distance from those not living in their household. Each state and territory had the power to decide on and enforce their own restrictions, with several states easing restrictions on organized sport and recreation facilities in early May due to no or low recorded cases of COVID-19 (see Figure 1).

Figure 1. Australian federal- and state-imposed restrictions to stop the spread of COVID-19. ACT: Australian Capital Territory; AUS: Australia; NSW: New South Wales; NT: Northern Territory; p: people; QLD: Queensland; SA: South Australia; TAS: Tasmania; VIC: Victoria; WA: Western Australia; WHO: World Health Organization.



Sample

Individuals living in Australia were recruited via social media advertising (eg, Facebook and Instagram), researcher and stakeholder organization networks, and snowballing techniques, wherein participants were asked to share the study information with others they knew (eg, word of mouth). Of 14,764 individuals who clicked on the link to the survey, 6474 provided informed consent (43.8% response rate), of which 4079 (63.0%) were adolescents (13-17 years) and 2395 (37.0%) were adults (18-75 years).

Measures

Demographic Characteristics

Participants reported their age (years), sex (male, female, other, or prefer not to say), usual daily responsibilities (paid employment status, home duties or carer responsibilities, and student status; yes or no), number of people living in the household, whether the English language was spoken at home (yes or no), and state or territory of residence.

Adherence to the Physical Activity Guidelines

Adherence to physical activity guidelines was assessed using valid and reliable survey items [17-19]. Participants were asked to indicate how many days per week (none, 1 day/week, 2 days/week, 3 days/week, 4 days/week, 5 days/week, 6 days/week, or 7 days/week) they performed MVPA during a usual week over the past month (ie, April or May 2020) and in February 2020 (ie, pre-COVID-19 restrictions) for 30 minutes (adults aged ≥18 years) or 60 minutes (adolescents aged 13-17 years). Adult responses were dichotomized at 5 or more days per week (ie, ≥150 minutes/week), and adolescent responses were dichotomized at 7 days per week (ie, ≥60 minutes/day). Participants were also asked whether they performed (yes or no) MSE during a usual week (at home over the past month [April or May] or at home or at a gym during February); those who said *yes* were then asked to report the number of times per week (ie, frequency) that they performed MSE during a usual week. This was used to determine adherence to the MSE guidelines (adolescents: ≥3 times/week; adults: ≥2 times/week). Adherence to the combined guidelines was also calculated for the past month (April or May) and February.

Digital Platforms for Physical Activity

Participants were asked to respond to the question "Are you currently doing any form of sports or physical activities using online or digital platforms to assist or guide your activity?" (yes or no). A comprehensive list of digital platforms was identified in collaboration with key stakeholders. Participants who responded *yes* were asked to report the frequency (number of times per week) and duration (total minutes per week) spent using each of six types of online or digital platforms: streaming services for exercise (eg, YouTube, Instagram, and Facebook); subscriber fitness programs, via an app or online (eg, Centr and MyFitnessPal); facilitated online live or recorded classes, via platforms such as Zoom (eg, dance, sport training, and fitness class); sport- or activity-specific apps designed by sporting organizations for participants to keep up their skills (eg, TeamBuildr); active electronic games (eg, Xbox Kinect); and/or online or digital training or racing platforms (eg, Zwift, FullGaz, and Rouvy).

Statistical Analysis

Demographic characteristics, physical activity, and digital platform use were presented descriptively. Demographic differences between those who used digital platforms (ie, users)

compared to those who did not (ie, nonusers) were calculated using chi-square tests and *t* tests. Unadjusted logistic regression models were used to identify associations between sample characteristics (ie, age, sex, English-speaking household, number of people in household, employment status, home duties or carer responsibilities, and student status) and MVPA, MSE, and combined physical activity guideline adherence. Adjusted multilevel logistic regression models were then run to examine associations between the use of digital platforms and adherence with MVPA, MSE, and both MVPA and MSE guidelines. All multilevel models accounted for clustering by state or territory of residence and were adjusted for sample characteristics found to be significant in univariate models (see Multimedia Appendix 1) and guideline adherence during February 2020. Analyses were stratified by age group (ie, adults and adolescents). All analyses were performed using Stata v16 (StataCorp LLC).

Results

Table 1 presents the sample characteristics for adult and adolescent participants. In total, 1188 adults and 963 adolescents provided complete physical activity guideline adherence and digital platform data and were included in analyses. Among adults, the mean age was 37.4 (SD 15.1) years, the majority were female, and two-thirds had a tertiary degree. In the past month, 33.0% (392/1188), 37.3% (443/1188), and 17.7% (210/1188) of adults met the MVPA, MSE, and both guidelines, respectively, and 39.5% (469/1188) used online or digital platforms to assist or guide their physical activity. Among those who used digital platforms for physical activity, the median frequency was 4 (IQR 2-6) times per week and the median duration was 105 (IQR 60-180) minutes per week. Among adolescents, the mean age was 16.2 (SD 1.2) years, and more than two-thirds were female. In the past month, 7.2% (69/963), 28.1% (271/963), and 3.6% (35/963) of adolescents met the MVPA, MSE, and both guidelines, respectively, and 26.5% (255/963) used online or digital platforms to guide or assist their physical activity. Among those who used digital platforms for physical activity, the median frequency was 4 (IQR 3-7) times per week and the median duration was 120 (IQR 60-260) minutes per week.

Table 1. Sample characteristics.

Characteristic	Adults (n=1188)	Adolescents (n=963)
Age in years, mean (SD)	37.4 (15.1)	16.2 (1.2)
Sex (female), n (%)	980 (82.5)	685 (71.1)
Employment status (working), n (%)	625 (52.6)	248 (25.8)
Home duties or carer responsibilities (yes), n (%)	180 (15.2)	127 (13.2)
Student status (yes), n (%)	258 (21.7)	622 (64.6)
Number of people in household, mean (SD)	3.2 (1.4)	4.3 (1.3)
English-speaking household (yes), n (%)	1155 (97.2)	931 (96.7)
State or territory of residence, n (%)		
Australian Capital Territory	47 (4.0)	23 (2.4)
New South Wales	217 (18.3)	234 (24.3)
Northern Territory	9 (0.8)	4 (0.4)
Queensland	124 (10.4)	197 (20.5)
South Australia	79 (6.6)	71 (7.4)
Tasmania	50 (4.2)	47 (4.9)
Victoria	590 (49.7)	305 (31.7)
Western Australia	72 (6.1)	82 (8.5)

Table 2 presents the demographic characteristics of digital platform users and nonusers. Among adults, a higher percentage of users were female and working in paid employment, and a higher percentage of nonusers had home duties or carer responsibilities. Among adolescents, a higher percentage of users were female.

Table 2. Demographic characteristics of users and nonusers of digital platforms for physical activity.

Characteristic	Adults (n=1188)			Adolescents (n=963)		
	Users	Nonusers	*P* value	Users	Nonusers	*P* value
Number of participants out of total, n (%)	469 (39.5)	719 (60.5)	N/A[a]	255 (26.5)	708 (73.5)	N/A
Age in years, mean (SD)	36.3 (13.2)	38.0 (16.2)	.06	16.2 (1.3)	16.3 (1.2)	.22
Sex (female), n (%)	428 (91.3)	552 (76.8)	<.001	213 (83.5)	472 (66.7)	<.001
Employment status (working), n (%)	281 (59.9)	344 (47.8)	<.001	63 (24.7)	185 (26.1)	.66
Home duties or carer responsibilities (yes), n (%)	56 (11.9)	124 (17.3)	.01	40 (15.7)	87 (12.3)	.17
Student status (yes), n (%)	89 (19.0)	169 (23.5)	.06	166 (65.1)	456 (64.4)	.84
English-speaking household (yes), n (%)	453 (96.6)	702 (97.6)	.28	245 (96.1)	686 (96.9)	.53
Number of people in household, mean (SD)	3.1 (1.3)	3.2 (1.5)	.56	4.4 (1.3)	4.3 (1.3)	.13

[a]N/A: not applicable; *P* values were not calculated for this item.

Among those who had used digital platforms to guide or assist their physical activity, the most common were streaming services (adults: 197/469, 42.0%; adolescents: 102/255, 40.0%), facilitated online classes (adults: 144/469, 30.7%; adolescents: 77/255, 30.2%), and subscriber fitness programs (adults: 139/469, 29.6%; adolescents: 35/255, 13.7%) (see Figure 2). The types of digital platforms used were generally similar for adults and adolescents; however, proportionally fewer adolescents used subscriber fitness programs.

Figure 2. Use of each type of digital platform for physical activity among users.

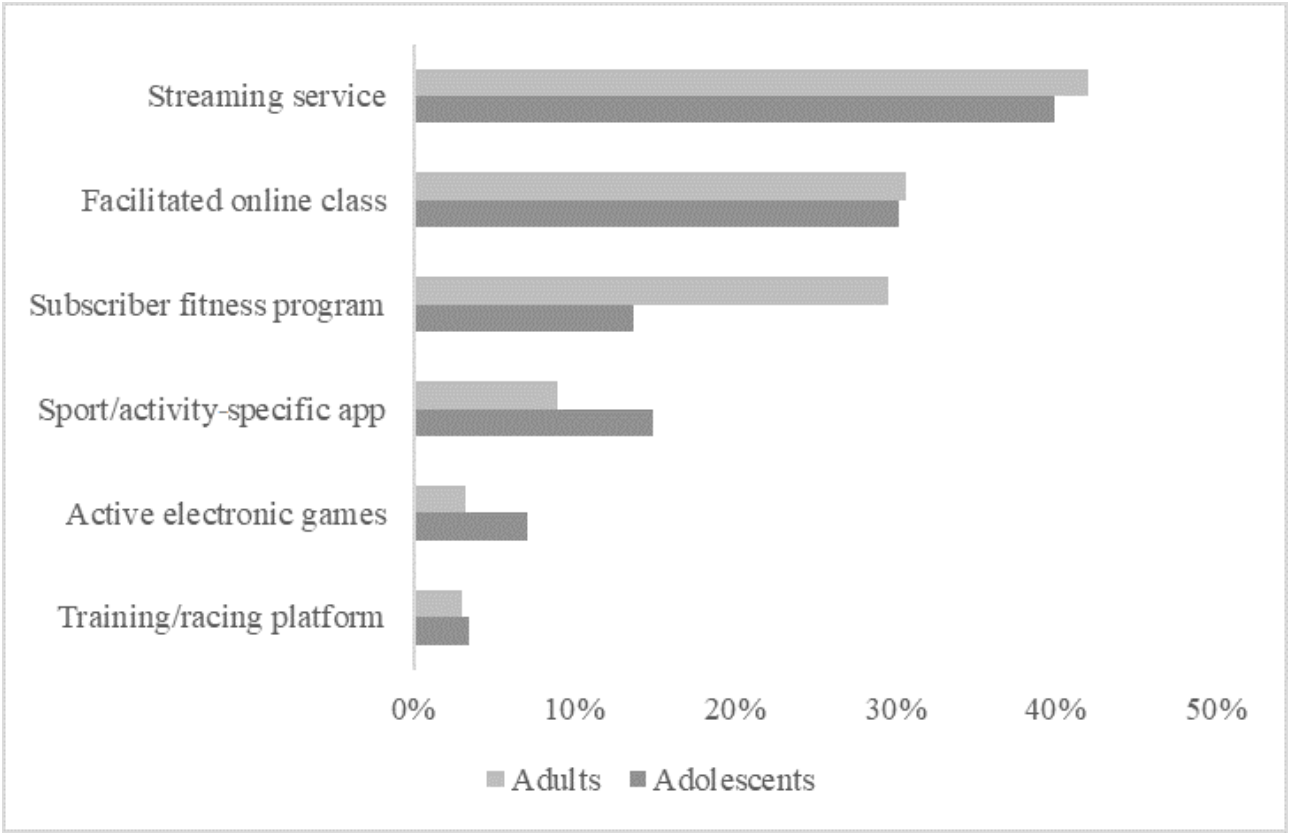


Table 3 presents the unadjusted and adjusted odds ratios for physical activity guideline adherence, accounting for state or territory of residence and adjusting for significant confounders (see Table S1 in Multimedia Appendix 1) and guideline adherence during February. In the adjusted models, adults who used digital platforms had 2 times the odds of meeting MVPA guidelines, over 3 times the odds of meeting MSE guidelines, and almost 3 times the odds of meeting the combined guidelines compared to nonusers. Adolescents who used digital platforms had more than 2 times the odds of meeting MVPA guidelines, over 3 times the odds of meeting MSE guidelines, and more than 4 times the odds of meeting the combined guidelines compared to nonusers.

Table 3. Adjusted odds ratios (ORs) for the associations of physical activity guideline adherence and use of digital platforms.

Model	Digital platform use for adults, OR (95% CI)[a]			Digital platform use for adolescents, OR (95% CI)[a]		
	MVPA[b]	MSE[c]	Combined	MVPA	MSE	Combined
Unadjusted model (reference: nonusers)	1.97 (1.54-2.52)	3.79 (2.96-4.87)	3.22 (2.36-4.39)	2.43 (1.47-3.99)	3.09 (2.27-4.20)	5.05 (2.50-10.18)
Adjusted model (reference: nonusers)	1.99 (1.49-2.66)[d]	3.34 (2.49- 4.47)[e]	2.73 (1.95-3.81)[f]	2.40 (1.32-4.35)[g]	3.07 (2.12-4.44)[h]	4.32 (2.09-8.95)[i]

[a]All accounted for clustering by state or territory and were adjusted for February guideline adherence.

[b]MVPA: moderate- to vigorous-intensity physical activity.

[c]MSE: muscle-strengthening exercise.

[d]Additionally adjusted for age, number of people living in household, home duties or carer responsibilities, and student status.

[e]Additionally adjusted for employment and home duties or carer responsibilities.

[f]Additionally adjusted for home duties or carer responsibilities.

[g]Additionally adjusted for student status.

[h]Additionally adjusted for sex and home duties or carer responsibilities.

[i]Additionally adjusted for home duties or carer responsibilities.

Discussion

Principal Findings

Findings from this study showed that digital platforms may play a critical role in supporting physical activity engagement during times when people have limited access to traditional settings or opportunities for physical activity outside the home. Individuals who used a digital platform were more likely to report achieving recommended levels of MVPA and MSE during April and May when strict stay-at-home orders were imposed for most of the nation.

Data sourced from the Global Digital Overview in January 2020 highlighted the ubiquitous use of digital platforms in Australia; 88% of the adult population had access to the internet, 71% used social media (eg, Facebook and Instagram), and 26% of internet users reported using health and fitness apps [20]. While there are no directly comparable data, data from our study suggest that a higher proportion of Australians used digital platforms for physical activity during April and May 2020 compared to 2018 [16], with a higher rate of use observed among adults compared with adolescents. In addition to the established use of mobile fitness programs [21], our findings suggest that people also used digital platforms for facilitated online live or recorded activity classes and for streaming services.

More females than males used digital platforms to guide or assist their physical activity, which is similar to previous research [22,23]. This may reflect the ability of digital platforms to support information sharing, self-monitoring, and internal accountability, which are often considered important for increasing physical activity motivation among females [24,25]. It may be that females participate more in instructor-led activities (eg, yoga, Pilates, and dance) [26], whereas males tend to engage in more organized sport and weight training [8], both of which are less adaptable to online delivery via digital platforms. Safety concerns when exercising alone outdoors or after dark and fear of judgement are known barriers to physical activity uptake by women [27,28]; this may have also informed women's decisions to use digital platforms to undertake physical activity in the home. Alternatives such as digital platforms may play an important role in ensuring that women achieve sufficient physical activity levels. Digital platforms can be used at any time of the day and offer convenience of use in the home, which may explain the larger proportion of adults working in paid employment using digital platforms for physical activity compared to those not working.

In this sample, 33% of adults reported meeting MVPA guidelines during the April and May stay-at-home period, which is considerably lower than the Australian average of 55% of 18- to 64-year-olds in 2017-2018 [8]. It should be noted that the MVPA measure in this study required adults to engage in 30 minutes per day of MVPA on at least 5 days per week, whereas the Australian Bureau of Statistics physical activity measure was based on a minimum total of 150 minutes of moderate-intensity physical activity, 75 minutes of vigorous-intensity physical activity, or an equivalent combination per week [8]. Among adolescents, just 7% met guidelines for MVPA, which was slightly lower than the Australian average of 10% of 15- to 17-year-olds [8]. These differences may have been due to the reduced ability to access traditional settings for physical activity, such as schools [29], work [30], and fitness and recreation facilities [31], or to participate in sport or active travel [32]. In contrast, the proportion of adults and adolescents in this sample who met the MSE guidelines during the April and May stay-at-home period was considerably higher than the Australian average [8]. MSE includes bodyweight activities that may not require specialized equipment or facilities; can be performed in a confined space, such as at home; and may have been promoted via digital platforms (eg, livestreams on YouTube) during the stay-at-home restriction phase.

Our results showed support for the use of digital platforms to engage in physical activity and MSE when access to traditional settings and facilities for physical activity was restricted. Adults and adolescents who used digital platforms to guide or assist physical activity were more likely to meet the MVPA, MSE, and combined physical activity guidelines compared to those

who did not use digital platforms. This is consistent with evidence from Germany [33] and the United States [22] that showed that adults who used physical activity and health apps engaged in more physical activity compared to those who did not. This study builds on the evidence that digital technologies can promote and support physical activity among adults and adolescents [13,34-39]. Streaming services were the most frequently used digital platform to guide or assist physical activity, so future studies should further explore how they can best support physical activity engagement for all demographic groups and how they can be used as a physical activity promotion tool. Streaming services are mostly free to use and, thus, present an opportunity for relevant government or nongovernment organizations to make use of this platform for education, instruction, and promotion of physical activity to the general public.

Overall, the findings highlight a willingness to engage with technology for MVPA and MSE when access to traditional settings or opportunities for physical activity outside the home are limited, particularly among females and working adults. Further research is needed to explore what motivated or discouraged people from using digital platforms during this period of reduced options for activity outside the home environment. Future work could also explore whether the use of digital platforms for physical activity replaced usual physical activity behaviors from before the restrictions or complemented other physical activities (eg, attending a fitness class in person) and how digital platforms can best support continued engagement in physical activity once restrictions are reduced.

Limitations

The large sample size and measures of MVPA, MSE, and combined guideline adherence are strengths of this study. However, the majority of participants were female and English speaking; in addition, 50% of adults and 32% of adolescents were from one state—Victoria—and, thus, are not representative of the wider Australian population. As the survey was completed online, English-language proficiency was required, which may have also reduced the generalizability of the findings to culturally and linguistically diverse populations. The measure used to capture MVPA guideline adherence in this study was valid and reliable [17-19]; however, it may not accurately capture individuals who engaged in shorter durations of vigorous-intensity physical activity yet still met the guidelines (eg, 30 minutes/day and 3 days/week). Participants were asked to report on MSE specifically in the home during April and May and at home or a gym in February; as such, this may not have captured all MSE performed (eg, in other locations). In addition, this study relied on self-report of physical activity and, as such, potential for recall bias must be acknowledged.

Conclusions

In this study, fewer than half of the adults and one-third of adolescents reported using digital platforms to assist or guide their physical activity during the COVID-19 stay-at-home period in April and May 2020. Both adults and adolescents who used digital platforms for physical activity were more likely to meet the MVPA, MSE, and combined physical activity guidelines compared to those who did not use digital platforms. This suggests that digital platforms can play a critical role in supporting physical activity engagement. There is a need for future research to understand sustained use, gender preferences, and motivations for the use of digital platforms to guide or assist physical activity, in particular via streaming services, given their popularity during COVID-19.

Acknowledgments

The authors thank all of the participants in OL@H. OL@H was supported by Deakin University Social and Early Emotional Development funding. JV and NR are each supported by an Australian National Heart Foundation Future Leader Fellowship (JV: ID 101928; NR: ID 101895). JS is supported by a National Health and Medical Research Council Investigator Grant (APP 1176885). RU is supported by an Alfred Deakin Postdoctoral Research Fellowship. SC is supported by an Australian Government Research Training Program Scholarship.

Authors' Contributions

KP, LA, RU, NDR, HB, JS, AT, SS, and JV were involved in the original design and conceptualization of the study. KP, LA, and SC were involved in the data collection. KP conducted the statistical analysis and interpretation of the statistics and was primarily responsible for drafting the manuscript. All authors reviewed the manuscript critically, contributed to the intellectual content, and approved the final version.

Conflicts of Interest

NDR and JS declare involvement in a start-up technological company. Other authors have no conflicts to declare.

Multimedia Appendix 1
Unadjusted odds ratios (95% CI) for the associations of sample characteristics and physical activity guideline adherence in April and May 2020.
[DOCX File , 22 KB - jmir_v23i2e23389_app1.docx]

References

1. Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016. Geneva, Switzerland: World Health Organization; 2018. URL: https://www.who.int/healthinfo/global_burden_disease/estimates/en/ [accessed 2020-06-29]
2. Strain T, Brage S, Sharp SJ, Richards J, Tainio M, Ding D, et al. Use of the prevented fraction for the population to determine deaths averted by existing prevalence of physical activity: A descriptive study. Lancet Glob Health 2020 Jul;8(7):e920-e930 [FREE Full text] [doi: 10.1016/S2214-109X(20)30211-4] [Medline: 32562648]
3. Global Recommendations on Physical Activity for Health. Geneva, Switzerland: World Health Organization; 2010. URL: https://www.who.int/dietphysicalactivity/publications/9789241599979/en/ [accessed 2020-06-19]
4. Australian 24-Hour Movement Guidelines for Children and Young People (5 to 17 Years): An Integration of Physical Activity, Sedentary Behaviour, and Sleep. Canberra, Australia: Australian Government Department of Health; 2019. URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/AC46EFBE3130BC97CA2583D00023E7B0/$File/Australian%2024-Hour%20Movement%20Guidelines%20for%20Children%20and%20Young%20People%205-17%20years.pdf [accessed 2020-07-11]
5. Australia's Physical Activity and Sedentary Behaviour Guidelines for Adults (18-64 Years). Canberra, Australia: Australian Government Department of Health; 2019. URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/health-pubhlth-strateg-phys-act-guidelines#npa1864 [accessed 2020-07-22]
6. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: A pooled analysis of 298 population-based surveys with 1.6 million participants. Lancet Child Adolesc Health 2020 Jan;4(1):23-35 [FREE Full text] [doi: 10.1016/S2352-4642(19)30323-2] [Medline: 31761562]
7. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: A pooled analysis of 358 population-based surveys with 1.9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
8. National Health Survey: First results. Australian Bureau of Statistics. 2019. URL: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4364.0.55.001~2017-18~Main%20Features~Physical%20activity~115 [accessed 2020-07-10]
9. Douglas M, Katikireddi S, Taulbut M, McKee M, McCartney G. Mitigating the wider health effects of COVID-19 pandemic response. BMJ 2020 Apr 27;369:m1557 [FREE Full text] [doi: 10.1136/bmj.m1557] [Medline: 32341002]
10. Wackerhage H, Everett R, Krüger K, Murgia M, Simon P, Gehlert S, et al. Sport, exercise and COVID-19, the disease caused by the SARS-CoV-2 coronavirus. Dtsch Z Sportmed 2020 May 01;71(5):E1-E12 [FREE Full text] [doi: 10.5960/dzsm.2020.441]
11. Ding D, Del Pozo Cruz B, Green MA, Bauman AE. Is the COVID-19 lockdown nudging people to be more active: A big data analysis. Br J Sports Med 2020 Oct;54(20):1183-1184. [doi: 10.1136/bjsports-2020-102575] [Medline: 32605932]
12. Sebire K. The coronavirus lockdown is forcing us to view 'screen time' differently. That's a good thing. The Conversation. 2020 Apr 12. URL: https://theconversation.com/the-coronavirus-lockdown-is-forcing-us-to-view-screen-time-differently-thats-a-good-thing-135641 [accessed 2020-05-20]
13. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
14. Stockwell S, Schofield P, Fisher A, Firth J, Jackson SE, Stubbs B, et al. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: A systematic review and meta-analysis. Exp Gerontol 2019 Jun;120:68-87. [doi: 10.1016/j.exger.2019.02.020] [Medline: 30836130]
15. Baranowski T. Exergaming: Hope for future physical activity? Or blight on mankind? J Sport Health Sci 2017 Mar;6(1):44-46 [FREE Full text] [doi: 10.1016/j.jshs.2016.11.006] [Medline: 30356568]
16. AusPlay results. Sport Australia. 2019. URL: https://www.clearinghouseforsport.gov.au/research/ausplay/results#previous_data_releases [accessed 2021-01-27]
17. Milton K, Bull FC, Bauman A. Reliability and validity testing of a single-item physical activity measure. Br J Sports Med 2011 Mar;45(3):203-208. [doi: 10.1136/bjsm.2009.068395] [Medline: 20484314]
18. Ridgers ND, Timperio A, Crawford D, Salmon J. Validity of a brief self-report instrument for assessing compliance with physical activity guidelines amongst adolescents. J Sci Med Sport 2012 Mar;15(2):136-141. [doi: 10.1016/j.jsams.2011.09.003] [Medline: 22051688]
19. Bennie JA, Lee D, Khan A, Wiesner GH, Bauman AE, Stamatakis E, et al. Muscle-strengthening exercise among 397,423 US adults: Prevalence, correlates, and associations with health conditions. Am J Prev Med 2018 Dec;55(6):864-874. [doi: 10.1016/j.amepre.2018.07.022] [Medline: 30458949]
20. Kemp S. Digital 2020: Global Digital Overview. New York, NY: We Are Social and Hootsuite; 2020. URL: https://wearesocial.com/digital-2020 [accessed 2020-07-15]
21. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: A systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]

22. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
23. Elavsky S, Smahel D, Machackova H. Who are mobile app users from healthy lifestyle websites? Analysis of patterns of app use and user characteristics. Transl Behav Med 2017 Dec;7(4):891-901 [FREE Full text] [doi: 10.1007/s13142-017-0525-x] [Medline: 28929368]
24. Lindgren T, Hooper J, Fukuoka Y. Perceptions and experiences of women participating in a digital technology-based physical activity intervention (the mPED Trial): Qualitative study. JMIR Public Health Surveill 2019 Dec 20;5(4):e13570 [FREE Full text] [doi: 10.2196/13570] [Medline: 31859677]
25. Carter A, Alexander AC. A qualitative exploration of womens' experiences who belong to a "fitness community". Am J Health Educ 2019 Dec 03;51(1):22-30. [doi: 10.1080/19325037.2019.1687365]
26. Vergeer I, Bennie JA, Charity MJ, Harvey JT, van Uffelen JGZ, Biddle SJH, et al. Participation trends in holistic movement practices: A 10-year comparison of yoga/Pilates and t'ai chi/qigong use among a national sample of 195,926 Australians. BMC Complement Altern Med 2017 Jun 06;17(1):296 [FREE Full text] [doi: 10.1186/s12906-017-1800-6] [Medline: 28587599]
27. Kilgour L, Parker A. Gender, physical activity and fear: Women, exercise and the great outdoors. Qual Res Sport Exerc Health 2013 Mar;5(1):43-57. [doi: 10.1080/2159676x.2012.718619]
28. Roper E. Concerns for personal safety among female recreational runners. Women Sport Phys Act J 2016;24(2):91-98. [doi: 10.1123/wspaj.2015-0013]
29. Erwin HE, Ickes M, Ahn S, Fedewa A. Impact of recess interventions on children's physical activity: A meta-analysis. Am J Health Promot 2014;28(3):159-167. [doi: 10.4278/ajhp.120926-LIT-470] [Medline: 23875990]
30. Bauman A, Ma G, Cuevas F, Omar Z, Waqanivalu T, Phongsavan P, Equity and Non-communicable Disease Risk Factors Project Collaborative Group. Cross-national comparisons of socioeconomic differences in the prevalence of leisure-time and occupational physical activity, and active commuting in six Asia-Pacific countries. J Epidemiol Community Health 2011 Jan;65(1):35-43. [doi: 10.1136/jech.2008.086710] [Medline: 20943821]
31. Move to be Well: The Global Economy of Physical Activity. Miami, FL: Global Wellness Institute; 2019 Oct. URL: https://globalwellnessinstitute.org/wp-content/uploads/2019/10/2019-Physical-Activity-Economy-FINAL-NEW-101019.pdf [accessed 2020-07-11]
32. Panik RT, Morris EA, Voulgaris CT. Does walking and bicycling more mean exercising less? Evidence from the US and the Netherlands. J Transp Health 2019 Dec;15:100590. [doi: 10.1016/j.jth.2019.100590]
33. Ernsting C, Dombrowski SU, Oedekoven M, O'Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: A population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
34. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: Behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
35. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: A systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
36. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: A systematic review and meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
37. Lee AM, Chavez S, Bian J, Thompson LA, Gurka MJ, Williamson VG, et al. Efficacy and effectiveness of mobile health technologies for facilitating physical activity in adolescents: Scoping review. JMIR Mhealth Uhealth 2019 Feb 12;7(2):e11847 [FREE Full text] [doi: 10.2196/11847] [Medline: 30747716]
38. Ferrer DA, Ellis R. A review of physical activity interventions delivered via Facebook. J Phys Act Health 2017 Oct 01;14(10):823-833. [doi: 10.1123/jpah.2016-0534] [Medline: 28556677]
39. Petersen JM, Prichard I, Kemps E. A comparison of physical activity mobile apps with and without existing web-based social networking platforms: Systematic review. J Med Internet Res 2019 Aug 16;21(8):e12687 [FREE Full text] [doi: 10.2196/12687] [Medline: 31420956]

Abbreviations

MSE: muscle-strengthening exercise
MVPA: moderate- to vigorous-intensity physical activity
OL@H: Our Life at Home study

Edited by G Eysenbach; submitted 11.08.20; peer-reviewed by J Bianchi-Hayes, A Davies; comments to author 07.10.20; revised version received 12.10.20; accepted 14.12.20; published 01.02.21.

Please cite as:
Parker K, Uddin R, Ridgers ND, Brown H, Veitch J, Salmon J, Timperio A, Sahlqvist S, Cassar S, Toffoletti K, Maddison R, Arundell L
The Use of Digital Platforms for Adults' and Adolescents' Physical Activity During the COVID-19 Pandemic (Our Life at Home): Survey Study
J Med Internet Res 2021;23(2):e23389
URL: https://www.jmir.org/2021/2/e23389
doi:10.2196/23389
PMID:33481759

Original Paper

Physical Activity Behavior Before, During, and After COVID-19 Restrictions: Longitudinal Smartphone-Tracking Study of Adults in the United Kingdom

Hannah McCarthy[1], MA; Henry W W Potts[2], BA, MSc, PhD; Abigail Fisher[3], BSc, PhD

[1]University College London, London, United Kingdom

[2]Institute of Health Informatics, University College London, London, United Kingdom

[3]Department of Behavioural Science and Health, University College London, London, United Kingdom

Corresponding Author:
Henry W W Potts, BA, MSc, PhD
Institute of Health Informatics
University College London
222 Euston Road
London, NW1 2DA
United Kingdom
Phone: 44 020 3549 5969
Email: h.potts@ucl.ac.uk

Abstract

Background: The COVID-19 pandemic led to the implementation of worldwide restrictive measures to reduce social contact and viral spread. These measures have been reported to have a negative effect on physical activity (PA). Studies of PA during the pandemic have primarily used self-reported data. The single academic study that used tracked data did not report on demographics.

Objective: This study aimed to explore patterns of smartphone-tracked activity before, during, and immediately after lockdown in the United Kingdom, and examine differences by sociodemographic characteristics and prior levels of PA.

Methods: Tracked longitudinal weekly minutes of PA were captured using the BetterPoints smartphone app between January and June 2020. Data were plotted by week, demographics, and activity levels at baseline. Nonparametric tests of difference were used to assess mean and median weekly minutes of activity at significant points before and during the lockdown, and as the lockdown was eased. Changes over time by demographics (age, gender, Index of Multiple Deprivation, baseline activity levels) were examined using generalized estimating equations (GEEs).

Results: There were 5395 users with a mean age of 41 years (SD 12) and 61% (n=3274) were female. At baseline, 26% (n=1422) of users were inactive, 23% (n=1240) were fairly active, and 51% (n=2733) were active. There was a relatively even spread across deprivation deciles (31% [n=1693] in the least deprived deciles and 23% in the most [n=1261]). We found significant changes in PA from the week before the first case of COVID-19 was announced (baseline) to the week that social distancing restrictions were relaxed (Friedman test: χ^2_2=2331, P<.001). By the first full week of lockdown, the median change in PA was 57 minutes less than baseline. This represents a 37% reduction in weekly minutes of PA. Overall, 63% of people decreased their level of activity between baseline and the first week of COVID-19 restrictions. Younger people showed more PA before lockdown but the least PA after lockdown. In contrast, those aged >65 years appeared to remain more active throughout and increased their activity levels as soon as lockdown was eased. Levels of PA among those classed as active at baseline showed a larger drop compared with those considered to be fairly active or inactive. Socioeconomic group and gender did not appear to be associated with changes in PA.

Conclusions: Our tracked PA data suggests a significant drop in PA during the United Kingdom's COVID-19 lockdown. Significant differences by age group and prior PA levels suggests that the government's response to COVID-19 needs to be sensitive to these individual differences and the government should react accordingly. Specifically, it should consider the impact on younger age groups, encourage everyone to increase their PA, and not assume that people will recover prior levels of PA on their own.

(J Med Internet Res 2021;23(2):e23701) doi:10.2196/23701

KEYWORDS

physical activity; mobile apps; apps; fitness trackers; mHealth; COVID-19; behavior; tracking; smartphone; pattern

Introduction

It is well established that sufficient physical activity (PA) is important for good health [1,2]. PA substantially reduces risks of common noncommunicable diseases including cardiovascular disease, diabetes, some cancers, and depression [1]. However, the COVID-19 pandemic resulted in worldwide implementations of restrictive measures to reduce social contact and viral spread. These varied by country, but have generally involved restrictions on nonessential movement and are likely to have had an impact on PA levels. Several papers have expressed concern about the negative consequences of reduced PA during these restrictions and the value of maintaining PA [3-6].

In the United Kingdom, throughout the COVID-19 lockdown, people were allowed to spend time outside for PA, with the exception of those who were "shielding" or those self-isolating due to COVID-19 exposure or symptoms. From March 23, 2020, to May 8 (Wales), May 11 (Scotland), or May 13 (England), individuals were allowed out for exercise once per day, respectively; subsequently, this changed to as often as was desired. This contrasted with more restrictive lockdowns seen in other countries, like Spain or France, where leaving home for exercise was not permitted. The inclusion of daily exercise in the United Kingdom lockdown guidance could account for the Sport England findings that 62% of 2000 people surveyed said they felt exercise was more important than before COVID-19 [7].

Evidence on the impact of pandemic restrictions on PA behavior varies and there are multiple methodological approaches available, each with strengths and limitations. Google Trends data showed that interest in exercise in April 2020 was higher than at any other time since records began. However, these data cannot tell us whether this increased interest was among those already habitually active or whether it translated into behavior change [8].

Cross-sectional surveys have been most commonly employed and have generally shown substantial declines in PA. The ECLB-COVID19 international survey gathered data from 1047 respondents from Africa (40%), Asia (36%), Europe (21%), and elsewhere (3%) April 6-11, 2020. They found substantial drops in PA in response to COVID-19 restrictions (eg, a 24% drop in the number of days/week of moderate-intensity PA and a 34% drop in the number of minutes of walking per day) [9]. Rogers and colleagues [10] conducted an online survey April 6-22, 2020, that found 25% of respondents reported doing less PA, while 12% reported doing more. Predictors of doing less PA were being female, not having access to a garden, having various pre-existing conditions, and expressing sentiments about personal or household risks. Older people (aged >70 years) were more likely to be doing the same intensity level of PA. People aged 20-34 were significantly more likely to have changed their PA levels to be either more or less intense than prior to the lockdown. More positive results were reported in a survey starting March 17, 2020, with 75% of respondents meeting PA guidelines. Meeting guidelines was associated with being female, being aged ≥65 years, having higher household income, and having had higher prior levels of PA, but negatively associated with prior physical symptoms [11].

All these surveys, which only captured the very early phases of the pandemic, used convenience sampling, using social media and snowballing approaches for recruitment. These approaches, while fast and low cost, can lead to sampling biases. In addition, it is very hard for individuals to accurately recall weekly minutes of PA even within recent timeframes (such as the last 7 days), let alone recalling PA prior to COVID-19 restrictions. Thus, these surveys are missing a reliable baseline measure. However, in recent years, many people have started routinely tracking their PA through the use of apps and wearables. This gives us a source of data that is longitudinal, predating the pandemic, and not reliant on self-report/recall.

One of the most widely used such technologies is the Fitbit; a blog by Fitbit described declines in PA in the week ending March 22, 2020 [12], including a 9% decline in step activity in the United Kingdom. All 20 countries studied showed declines, with the largest being a 38% decline in Spain, which had a more restrictive lockdown. By June 2020, Fitbit reported step count levels increasing but not yet back to the same level as last year, except among older women (aged 50-64 years), who surpassed the previous year's levels. Younger Fitbit users seemed to be making the smallest step count gains. Fitbit differentiates between steps and active minutes, which are defined as more "vigorous and intentional" activity that is important for heart health. Vigorous activity had increased during the lockdown, but appeared to be returning to the same levels as last year across all age groups [13]. Similar trends were seen among users of Garmin fitness trackers, showing a reduction in steps and an increase in more vigorous PA [14]. Neither data set has been described in academic reports.

The largest tracking study to date had walking data from 455,404 users of a smartphone app in 187 countries and found a 27% decrease in steps between baseline and a month after the announcement of COVID-19 as a pandemic [15]. No demographic data were available, meaning it was not possible to characterize users or explore sociodemographic patterns.

Therefore, the aims of this study were to explore patterns of tracked activity (ie, walking, running, and cycling) in the United Kingdom before, during, and immediately after the COVID-19 restrictions were in place and to explore variations by demographic characteristics.

Methods

Overview

Participants were individuals in the United Kingdom registered with BetterPoints. BetterPoints is a free, publicly available, smartphone-based program that offers rewards (points, lottery style tickets, and virtual rewards such as medals) for the amount of PA tracked per week. Points can ultimately be converted to

financial rewards (exact amounts and types are dependent on the program sponsor). Program sponsors include local government/councils, National Health Service trusts, Clinical Commissioning Groups, Development Corporations and, increasingly, large corporate entities. Further information about BetterPoints can be found on their website [16].

Individuals must be aged ≥14 years to register with the program. Registered users who had tracked any PA at all between January 22 and June 17, 2020, were included in the study. On registration with the app, users are asked to provide year of birth (used to derive approximate age), gender (male/female/other), and home post code (used to derive the Index of Multiple Deprivation [IMD] decile via the UK Government website [17]).

This study was approved by the UCL ethics committee (ID 401.001). When registering with the BetterPoints program, individuals agree and consent to the Terms and Conditions and Privacy Statement, including that their tracked data will be used to monitor patterns and that their "anonymized data may be shared with trusted non-BetterPoints entities to do research."

PA was tracked by the BetterPoints smartphone app. Users could track their activities via a menu where they select activity types such as walking, running, or cycling, or they could turn on automatic tracking. From March 2020, automatic tracking became the default. BetterPoints uses proprietary algorithms to combine data from the chipsets in the phone (motion sensors, accelerometers, built-in classifiers) with additional data pertaining to speed, global positioning system data, and various map data sources to classify activity types automatically. The BetterPoints system records 0 if no valid activity is tracked. This is designed to avoid categorizing small amounts of movement (eg, walking around the house). The person must move over a distance, at a certain speed and acceleration, for the movement to qualify as walking, running, or cycling. To run automatic tracking, the smartphone must have a motion co-processor or accelerometer that monitors movement. Most current smartphones support this automatic tracking, including the Apple iPhone 5S (iOS) or above and nearly all Android-based phones. Smartphone sensors in iOS and Android phones have been shown to provide valid estimates of PA in naturalistic settings compared to ActiGraph [18] and pedometers [19].

The BetterPoints app displays data in a dashboard view of total weekly minutes of PA, which incorporates time spent walking, running, and cycling.

Analysis

This was a retrospective study design using existing data in the context of a constantly evolving pandemic response, so pragmatic decisions had to be made, including about how to define a baseline period and which follow-up measurement periods should be included in some analyses. Key dates in the United Kingdom's COVID-19 pandemic response were chosen, within the context of the data having a resolution of one week and attempting to reduce the number of unnecessary post hoc statistical tests. A summary of our measurement dates is provided in Table 1. Our weeks run from Wednesday to Tuesday because our data set began January 1, 2020, which was a Wednesday. The week commencing January 22, 2020, was selected as the baseline. This was the week before the first case of COVID-19 was reported in the United Kingdom.

Table 1. Baseline and follow-up measurement dates for analysis.

Date (week commencing)	Significant events
January 22	Study baseline (the week before the first COVID-19 case in the United Kingdom was announced)
March 11	Nonessential travel banned and social distancing introduced
March 18	The lockdown began on March 23
March 25	First full week of the lockdown
May 13	The lockdown was relaxed (multiple excursions for exercise allowed)
June 17	First full week with nonessential shops reopened (reopened on June 15)

Demographics were summarized using descriptive statistics. Participants were grouped into Sport England Active Lives' categories of active (≥150 minutes of PA per week), fairly active (30-149 minutes), and inactive (0-29 minutes), according to the minutes of activity participants engaged in during the baseline week (commencing January 22, 2020) [20].

PA data were highly skewed, with many zero values (39% of all data values), so data were analyzed in three ways. First, median and mean PA were plotted by week, overall and median by demographics and activity levels at baseline. Next, nonparametric tests were used to compare PA over time. We performed a Friedman test to determine if there were significant differences in activity over time, then Wilcoxon tests were conducted to compare key follow-up weeks to baseline. Change in PA was calculated by computing baseline minus follow-up week. The mean, median, and interquartile range of this change score were then calculated. The change was also expressed as a percentage of baseline. Change categories were used to describe the proportion of people who had decreased, increased, or maintained their activity levels from baseline to the first full week of COVID-19 restrictions. These analyses were conducted in SPSS (Version 26, IBM Corp).

Finally, parametric tests were performed; specifically, generalized estimating equations (GEEs), a further generalization of the generalized linear model (GLM), were used to take into account the variance structure of the outcome data being from the same individuals over multiple weeks. We wanted to use a negative binomial regression, appropriate for the highly skewed data [21], but the GEE models failed to converge, presumably the result of a large number of zero values

and high correlations between each week and the next. Instead, additional change scores were created. We calculated 20 sets of values using the amount of PA in each week minus the previous week for that individual. This change score is more amenable to analysis, producing a symmetric distribution (skew=–0.1). However, it is very leptokurtic (kurtosis=17.9). We carried out a modified square root transformation as shown in equation 1:

$$\text{transformed value} = \text{sign}(x) \times \sqrt{(|x|)} \quad (1)$$

This reduces the length of the data distribution's tails and the kurtosis (now 3.3). We were then able to fit a linear regression GEE to the data. We used Huber-White sandwich robust variance estimators. These analyses were conducted in Stata (Version 11, StataCorp).

Results

Demographics

In total, 5395 users registered at least some activity each week from the week commencing January 22 to June 17, 2020, and were included in the current analysis. Participant characteristics are provided in Table 2. There were 130 missing values for gender data. In addition, 15 cases of approximate age >100 years were recoded as missing. Since there was no missing data for the primary outcome (PA) and <4% missing data for age or IMD, the decision was made not to impute these values. Users were on average 41 years old (SD 12; range 14-93 years). In addition, 61% of users identified as female, 37% as male, and 0.4% as other. There was a relatively even spread across deprivation deciles, with 31% falling in the least deprived deciles, and 23% in the most. At baseline, 26% of users were inactive, 23% were fairly active, and 51% were active.

Table 2. Participant characteristics and proportions of missing data (N=5395).

Characteristics	Values
Age (years), mean (SD)[a]	41.02 (12.2)
Age categories (years), n (%)	
14-24	463 (8.6)
25-34	1210 (22.4)
35-44	1554 (28.8)
45-54	1246 (23.1)
55-64	567 (10.5)
≥65	168 (3.1)
Gender, n (%)	
Male	1971 (36.5)
Female	3274 (60.7)
Other	20 (0.4)
Index of Multiple Deprivation, n (%)	
1-3 (most deprived)	1261 (23.4)
4-7	2240 (41.5)
8-10 (least deprived)	1693 (31.4)
Missing	201 (3.7)
Baseline physical activity, n (%)	
Inactive	1422 (26.4)
Fairly active	1240 (23.0)
Active	2733 (50.7)

[a]There were 187 missing values.

Plots and Nonparametric Tests

Plots of PA over the study period are shown in Figures 1-5. PA started to decline the week commencing March 11, 2020, when nonessential travel and social distancing measures were introduced. This decline continued until the first week of full lockdown, after which it appeared to remain fairly static throughout most of the lockdown, apart from a blip for the week commencing June 3, which had inclement weather.

Figure 1. Weekly minutes of physical activity (mean and median). Important dates are marked on the x-axis as follows. (A) January 22: Baseline, week before first COVID-19 case announced in the United Kingdom. (B) March 11: Social distancing measures introduced. (C) March 18: Lockdown begins. (D) March 25: First full week of lockdown. (E) May 13: Lockdown measures relaxed. (F) June 17: Shops reopen.

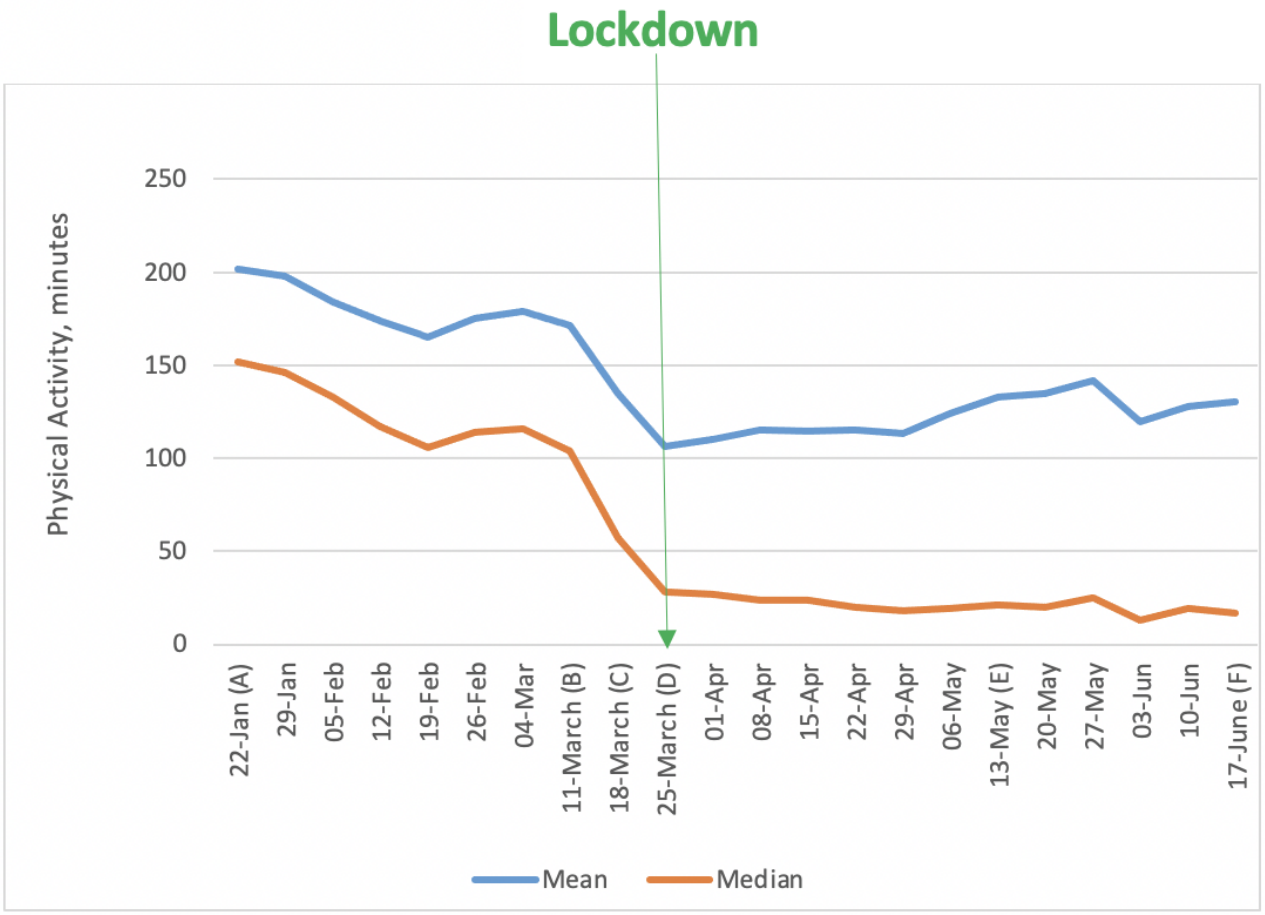


Figure 2. Weekly minutes (median) of activity in males and females (the "other" group was too small). Important dates are marked on the x-axis as follows. (A) January 22: Baseline, week before first COVID-19 case announced in the United Kingdom. (B) March 11: Social distancing measures introduced. (C) March 18: Lockdown begins. (D) March 25: First full week of lockdown. (E) May 13: Lockdown measures relaxed. (F) June 17: Shops reopen.

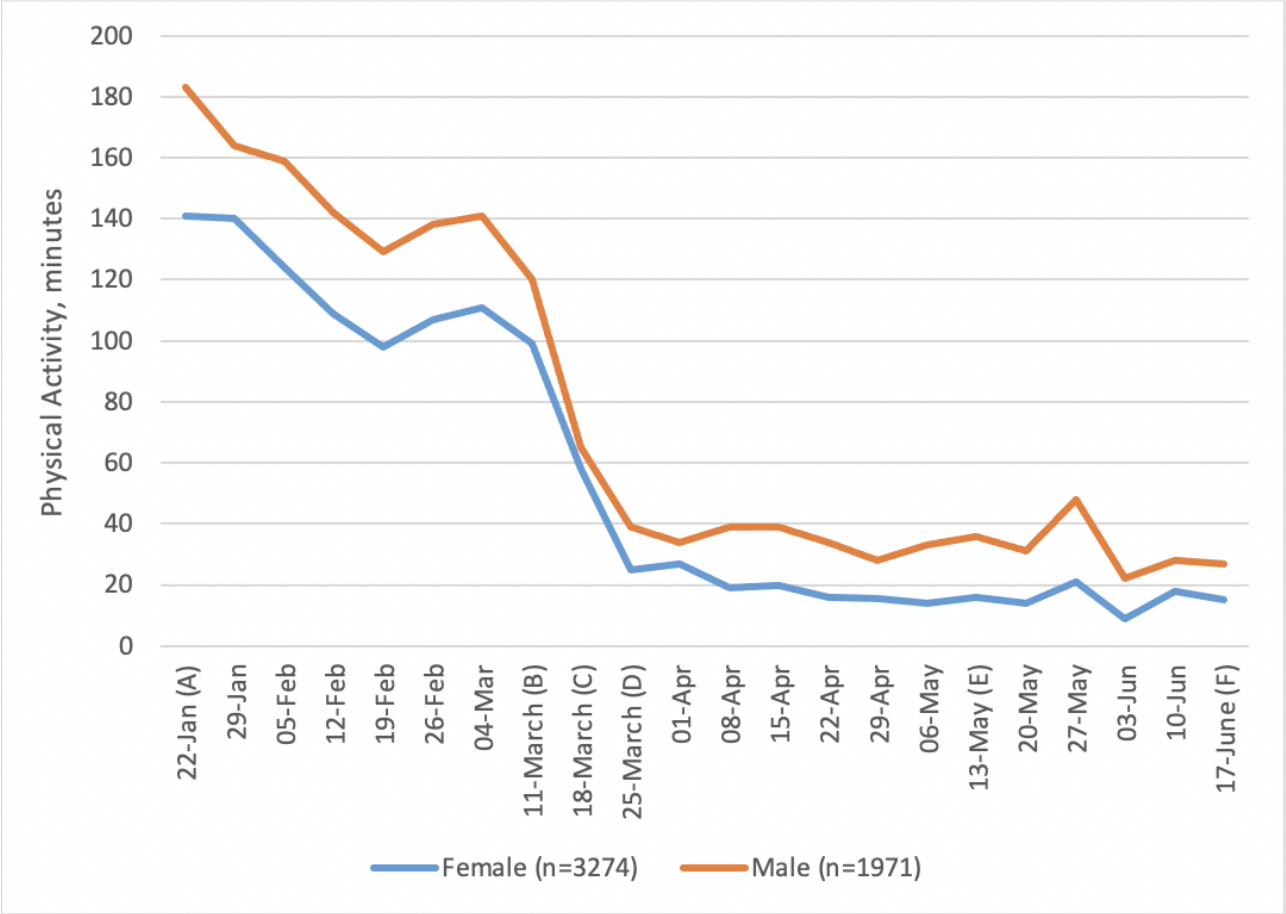


Figure 3. Weekly minutes (median) of activity by age. Important dates are marked on the x-axis as follows. (A) January 22: Baseline, week before first COVID-19 case announced in the United Kingdom. (B) March 11: Social distancing measures introduced. (C) March 18: Lockdown begins. (D) March 25: First full week of lockdown. (E) May 13: Lockdown measures relaxed. (F) June 17: Shops reopen.

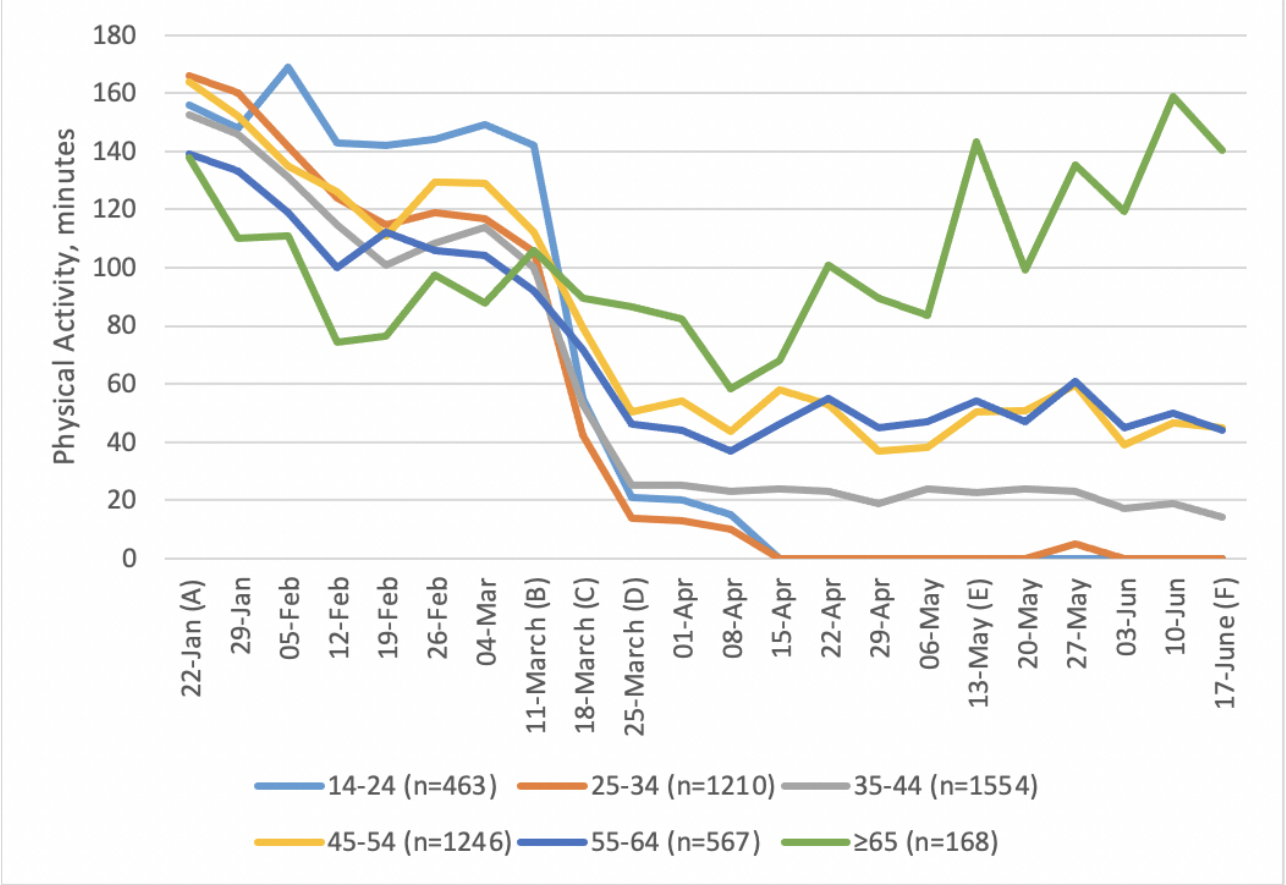


Figure 4. Weekly minutes (median) of activity by physical activity level at baseline. Important dates are marked on the x-axis as follows. (A) January 22: Baseline, week before first COVID-19 case announced in the United Kingdom. (B) March 11: Social distancing measures introduced. (C) March 18: Lockdown begins. (D) March 25: First full week of lockdown. (E) May 13: Lockdown measures relaxed. (F) June 17: Shops reopen.

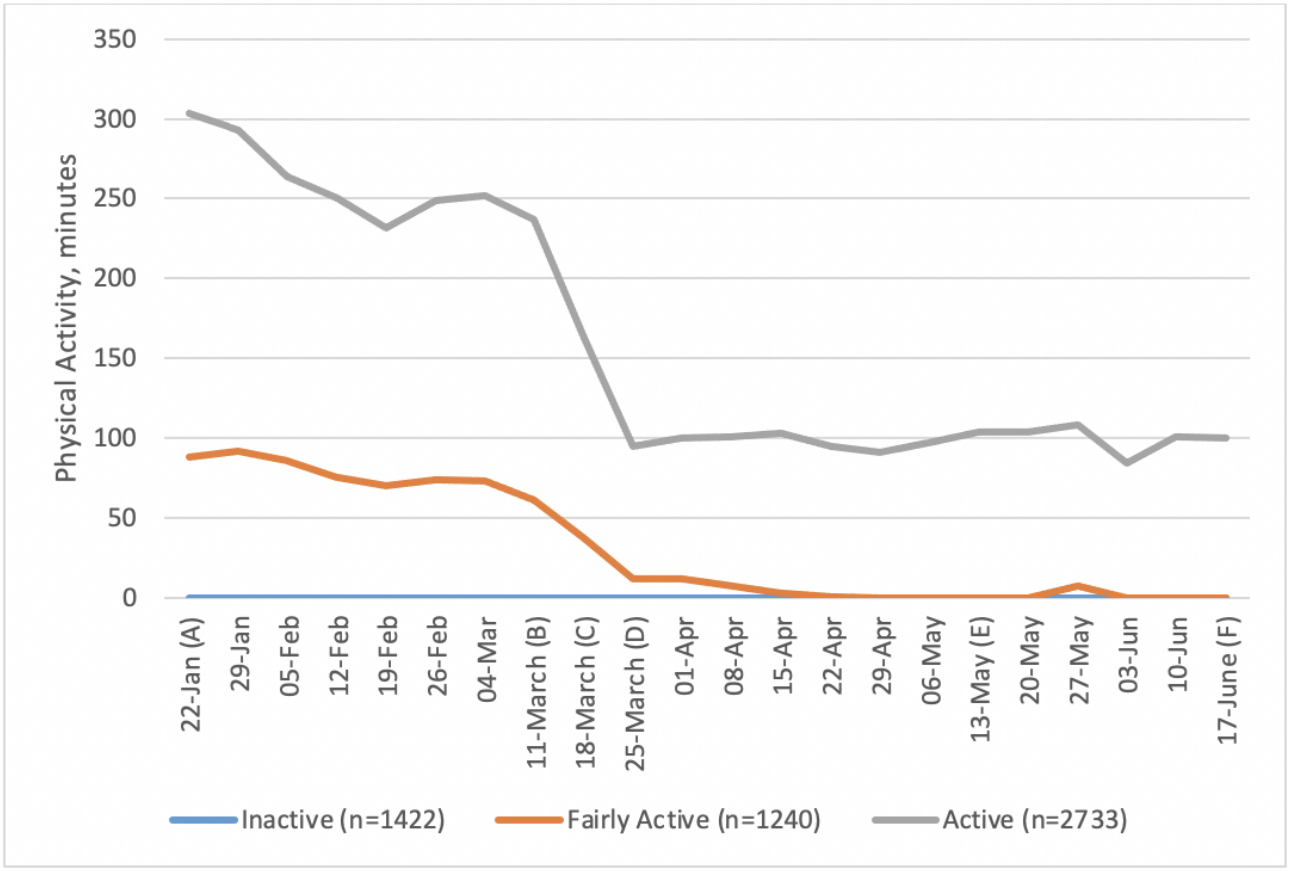
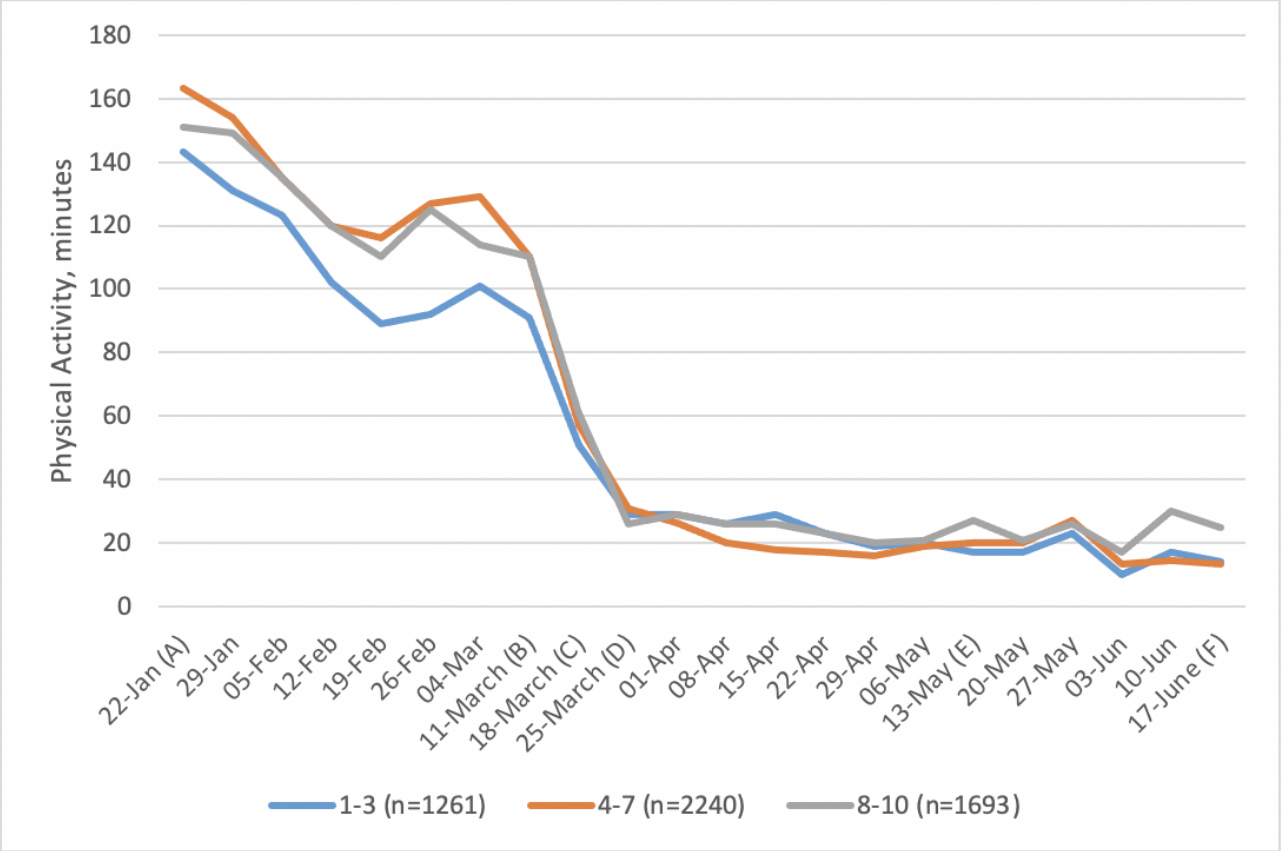


Figure 5. Weekly minutes (median) of activity by indices of multiple deprivation. Important dates are marked on the x-axis as follows. (A) January 22: Baseline, week before first COVID-19 case announced in the United Kingdom. (B) March 11: Social distancing measures introduced. (C) March 18: Lockdown begins. (D) March 25: First full week of lockdown. (E) May 13: Lockdown measures relaxed. (F) June 17: Shops reopen.



Median weekly minutes of PA and the results of statistical tests of difference are presented in Table 3. There were significant differences in PA between the week before the first case of COVID-19 was announced (baseline) and the week that social distancing restrictions were relaxed (Friedman test: χ^2_2=2331, P<.001).

Table 3. Summary of physical activity at baseline, through lockdown, and beyond (January 22-June 17, 2020; N=5395).

Time points	Physical activity (minutes), median (IQR)	Friedman test
Baseline	152 (20-306)	N/A[a]
Social distancing introduced (March 11)	104 (0-273)	N/A
Lockdown begins (March 18)	57 (0-209)	N/A
First full week of Lockdown (March 25)	28 (0-158)	N/A
Lockdown relaxed (May 13)	21 (0-199)	N/A
Shops reopen (June 17)	17 (0-197)	χ^2_2=2331, P<.001

[a]N/A: not applicable.

There were statistically significant changes in weekly minutes of tracked PA at all time points leading up to lockdown and in the weeks following the easing of lockdown measures (Table 4). There was a 2-minute reduction in activity between the baseline week (January 22, 2020) and the week nonessential travel and social distance restrictions were introduced. The week that the lockdown was announced, PA dropped by more than 30 minutes. By the following week (the first full week of lockdown restrictions), PA was down by 57 minutes compared with baseline. The drop in PA from baseline to the lockdown represents a 37% reduction in PA.

Table 4. Wilcoxon tests of change in physical activity from baseline to key weeks during the COVID-19 restrictions (January 22-June 17, 2020).

Key events (week commencing)	Mean change (minutes)	Percentage decrease from baseline	Median change (minutes)	Percentage decrease from baseline	*P* value
Social distancing introduced (March 11)	30	15	2	1	<.001
The lockdown begins (March 18)	67	33	32	21	<.001
First full week of the lockdown (March 25)	95	47	57	37	<.001
The lockdown is relaxed (May 13)	69	34	42	28	<.001
General shops reopen (June 17)	71	35	41	27	<.001

Overall, 63% of people decreased their level of activity between baseline and the first week of COVID-19 restrictions, 16% of people did not change their PA, and 21% increased their PA. The median change in PA was very similar for males and females (Figure 2). There appeared to be an effect of age, whereby younger people engaged in more PA before lockdown and the least amount of PA after lockdown (Figure 3). In contrast, those aged ≥65 years appeared to remain more active throughout and increased their activity levels as soon as the lockdown was eased.

Levels of PA among those classed as active at baseline showed a dramatic drop (Figure 4). Median levels of PA among those classed as fairly active at baseline also dropped but less dramatically. People who were inactive at baseline remained inactive throughout. Socioeconomic group did not appear to be associated with changes in PA (Figure 5).

GEE Models

A linear GEE was fitted to the data, with transformed week-on-week change in PA as the dependent variable. We explored the independent variables of gender, approximate age (as a continuous variable), deprivation decile, baseline activity level (inactive, fairly active, or active), and week (as a categorical variable). Examination of the data suggested an autoregressive correlation structure was appropriate. We fitted an AR(1) structure [22]. The resulting coefficients do not have a simple interpretation given the use of transformed data, but positive values indicate week-on-week increases in PA, while negative values indicate week-on-week decreases. Larger coefficients indicate larger changes. Sensitivity analyses were carried out using different assumptions for the correlation structure and with alternative modelling approaches. Similar results were found.

A basic model was fitted with just week as an independent variable, included as a categorical variable. This was statistically significant (χ^2_{20}=1759, *P*<.001). We then tested a model with the additional independent variables of age, gender, baseline activity, and deprivation index. This was statistically significant (χ^2_{24}=2562, *P*<.001). A likelihood ratio test showed improved fit from the additional variables (χ^2_4=895, *P*<.001). This indicated that older individuals showed more week-on-week increases in PA; that those who were most active showed more week-on-week decreases; and a pattern of increases and decreases over time matching Figure 1. All of these effects are independent of each other. There was no statistically significant relationship between week-on-week change and either gender or deprivation index, as shown in Table 5.

Table 5. Statistical analysis.

Variable	Coefficient	*P* value
Age (per decade)	0.08	<.001
Gender	–0.01	.50
Baseline activity (3 levels)	–0.38	<.001
Deprivation index	0.00	.60
Change from previous week to week n		
To week 2	0.84	<.001
To week 3	0.61	.001
To week 4	0.32	.07
To week 5	1.10	<.001
To week 6	0.54	.002
To week 7	0.33	.06
To week 8 (social distancing introduced)	–2.45	<.001
To week 9 (the lockdown begins)	–1.90	<.001
To week 10 (first full week of the lockdown)	0.55	.001
To week 11	0.62	<.001
To week 12	0.27	.10
To week 13	0.20	.20
To week 14	0.17	.30
To week 15	0.97	<.001
To week 16	0.94	<.001
To week 17 (the lockdown is relaxed)	0.28	.09
To week 18	0.83	<.001
To week 19	–1.40	<.001
To week 20	0.88	<.001
To week 21	0.50	.002
To week 22 (general shops reopen)	Baseline	Baseline

We additionally wanted to investigate whether the relationship with age varied over time. We simplified the time variable, dividing the weeks into three phases: weeks 1-7 (prelockdown), weeks 8-9 (the lockdown started), weeks 10-21 (the lockdown continued). A model using these phases as independent variables rather than individual weeks does not fit as well, but it fits well enough to allow the investigation of interaction effects. We tested the addition of two interaction terms for age in the "lockdown started" phase and the "lockdown continued" phase: these were statistically significant (likelihood ratio test, χ^2_2= 49, *P*<.001). This confirmed that older participants showed less decrease in PA when the lockdown began, and a greater increase in PA as the lockdown continued, confirming the interpretation of Figure 3.

Discussion

This longitudinal study of tracked PA, before, during, and after COVID-19 restrictions, showed large decreases in PA. Significant decreases in PA were observed at all time points from the week that social distancing measures were introduced, throughout the lockdown and the week measures were relaxed. PA was still significantly lower by the week commencing June 17, when nonessential shops reopened. The week that the lockdown was announced, median PA was down by 30 minutes and by the first full week of the lockdown, it had reduced by nearly an hour a week (57 minutes). This drop in PA represents a 37% reduction in individuals' weekly minutes of PA. Older people were significantly more likely to maintain and then increase their PA levels during the lockdown. Those who were most active to begin with showed the biggest falls in PA, but they had the furthest to fall. Although men showed more PA on average throughout, in our GEE model there was no effect of gender on the decline of PA. The deprivation index did not show any relationship with PA, although IMD is only an approximate measure of an individual's socioeconomic status, and the lack of evidence here may not translate into a lack of important socioeconomic effects.

Decreases in our study were substantially larger than the 9% drop observed from UK Fitbit step data [12]. Fitbit data ended

the week of March 22, which was just before the full lockdown restrictions were in place, and our data showed a 21% decrease in PA in the same week (week commencing March 18; our weeks ran Wednesday-Tuesday). Our smartphone-based measure may have overestimated the decline, since activity like incidental steps accumulated in the home or workplace would not have been captured. However, our study also showed greater decreases in PA than found in the large international tracking study by Tison and colleagues [15], who also using smartphone tracking. UK data from that study showed approximately a 10% decrease in mean steps from their prepandemic baseline level (mean daily steps from January 19 to March 11, 2020) to March 25, compared with a 37% decrease in tracked minutes in this study. It is not until the first week of April that percentage decreases in UK step counts converged with our findings, at around a 33% decrease from baseline, although our data included an estimate of time spent cycling as well as walking/running. The percentage drop in weekly PA in our findings is also in line with the international ECLB-COVID19 study, which found a reported 34% drop in the number of minutes of walking per day [9].

Overall, our findings are less optimistic than some studies of PA response to the pandemic that relied on data captured at the start of social restrictions. Increased interest in and intention to exercise have been reported elsewhere [8,20]. If such intentions existed in our sample, it appears they were only translated into action among older age groups. Those who were more active at baseline had the largest drops in PA and inactive people remained so throughout, contrary to findings that suggest a surge in PA during lockdown [10,11,23]. Perhaps of most concern are the fairly active group, who were close to doing recommended levels of PA for a number of weeks prior to lockdown and then dropped to doing no PA, with no sign of change as the lockdown was eased. We should not presume that these people will just automatically return to their prelockdown levels of exercise as restrictions are lifted. Measures to encourage this group back to prior levels of PA and continue to build on those efforts seem worthwhile.

In our study, older people showed less of a decrease in PA when the lockdown was introduced and recovered their PA levels during the lockdown faster than their younger counterparts. Concordant results were found by Fitbit [13] and by Smith and colleagues [11,24]. Our findings build on these prior results in suggesting that differences in maintained and increased PA in older age groups continued throughout the lockdown (and beyond). It is particularly interesting to note that approximately half of our sample of older adults were aged >70 years; that group had been urged by the UK Government not to go outside for even one daily bout of exercise. Although this is reassuring news about PA behavior in those aged >65 years, the large reduction in PA levels and failure to show any recovery for those below middle age is concerning. This picture suggests that policy interventions may be better focused on the long-term impact of restrictions on younger groups, which supports conclusions from recent research [11].

It was not possible to establish why we observed this pattern with age in our data. However, there is evidence that younger people are more worried about COVID-19, which may be a factor [25]. It is also feasible that factors like working/schooling from home had a stronger influence in younger age groups, but this is an area for future study.

Our study only tracked PA accumulated outdoors; some people may have substituted outdoor PA with indoor PA. BetterPoints users were given the opportunity to record activities such as “Be active at home” and “Try something new” during the lockdown, but analysis of this data was beyond the scope of the present analysis and may be included in future studies. The inclusion of live-streamed and prerecorded exercise classes in the app is currently being explored, while ways to routinely capture this type of activity and further surveying of users to understand whether they shifted to indoor exercise is underway. Wearables data suggest that while step counts may have reduced during the pandemic, vigorous activity may have increased, although these gains have not necessarily been maintained as we emerged from the lockdown [13]. This kind of activity may have continued among the previously active group but cannot be known from the current data.

While the approach of using app data gives us a good baseline and more valid measures of PA, we do have a sample bias as the results depend upon who uses BetterPoints. It may be that older adults who are using the BetterPoints app are a particular group who are more interested in exercise in the first place. Other limitations inherent in using data collected from smartphones include variability in how people carry and interact with their phones and variability in how many people track data on any given day. Although data on age, gender, and IMD were collected, information on other important demographics (like ethnicity) were not. In addition, although the app is free to use, smartphone ownership may be less common in some groups (eg, those with very low income). Only 79% of the UK population has a smartphone (2019 data), which means we cannot be sure how representative our data are of the UK population [26].

We have shown that tracked PA data suggests larger drops in PA during the United Kingdom’s COVID-19 lockdown than indicated previously. Significant differences by age group and prior PA levels suggests that blanket conclusions cannot be drawn about the impact of social distancing measures on population PA. The importance of better understanding in how to engage with and support different groups in tailored ways cannot be underestimated. Government response to COVID-19, particularly during the current situation where renewed outbreaks lead to local restrictions being imposed, needs to be sensitive to these individual differences and the government must react accordingly.

Authors' Contributions

All authors designed the study. HM handled data extraction and management. HM and HP did the analyses. HM wrote the first draft. All authors reviewed and edited the manuscript, revising it for intellectual content. All authors gave their final approval for publication and agreed to be accountable for all aspects of the work.

Conflicts of Interest

HM works for BetterPoints but is also studying for a doctorate at UCL. The study's other authors have no financial connection to BetterPoints. The company only shared anonymized data in line with privacy and data protection laws as approved by the UCL Ethics Committee and did not provide input into the study design, results, or analysis of the research. AF and HP have had a recent research project with Six to Start, who make apps promoting physical activity.

References

1. Warburton D, Bredin SSD. Health benefits of physical activity: a systematic review of current systematic reviews. Curr Opin Cardiol 2017 Sep;32(5):541-556. [doi: 10.1097/HCO.0000000000000437] [Medline: 28708630]
2. Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. BMJ 2019 Aug 21;366:4570 [FREE Full text] [doi: 10.1136/bmj.l4570] [Medline: 31434697]
3. Hammami A, Harrabi B, Mohr M, Krustrup P. Physical activity and coronavirus disease 2019 (COVID-19): specific recommendations for home-based physical training. Managing Sport and Leisure 2020 Apr 20;19:1-6. [doi: 10.1080/23750472.2020.1757494]
4. Jiménez-Pavón D, Carbonell-Baeza A, Lavie CJ. Physical exercise as therapy to fight against the mental and physical consequences of COVID-19 quarantine: Special focus in older people. Prog Cardiovasc Dis 2020;63(3):386-388 [FREE Full text] [doi: 10.1016/j.pcad.2020.03.009] [Medline: 32220590]
5. Lippi G, Henry BM, Sanchis-Gomar F. Physical inactivity and cardiovascular disease at the time of coronavirus disease 2019 (COVID-19). Eur J Prev Cardiol 2020 Jun;27(9):906-908 [FREE Full text] [doi: 10.1177/2047487320916823] [Medline: 32270698]
6. Rundle AG, Park Y, Herbstman JB, Kinsey EW, Wang YC. COVID-19-Related School Closings and Risk of Weight Gain Among Children. Obesity (Silver Spring) 2020 Jun 18;28(6):1008-1009 [FREE Full text] [doi: 10.1002/oby.22813] [Medline: 32227671]
7. Sport England. New exercise habits forming during coronavirus crisis. 2020. URL: https://www.sportengland.org/news/new-exercise-habits-forming-during-coronavirus-crisis [accessed 2020-12-23]
8. Ding D, Del Pozo Cruz B, Green MA, Bauman AE. Is the COVID-19 lockdown nudging people to be more active: a big data analysis. Br J Sports Med 2020 Oct 30;54(20):1183-1184. [doi: 10.1136/bjsports-2020-102575] [Medline: 32605932]
9. Ammar A, Brach M, Trabelsi K, Chtourou H, Boukhris O, Masmoudi L, et al. Effects of COVID-19 Home Confinement on Eating Behaviour and Physical Activity: Results of the ECLB-COVID19 International Online Survey. Nutrients 2020 May 28;12(6):1 [FREE Full text] [doi: 10.3390/nu12061583] [Medline: 32481594]
10. Rogers NT, Waterlow NR, Brindle H, Enria L, Eggo RM, Lees S, et al. Behavioral Change Towards Reduced Intensity Physical Activity Is Disproportionately Prevalent Among Adults With Serious Health Issues or Self-Perception of High Risk During the UK COVID-19 Lockdown. Front Public Health 2020;8:575091 [FREE Full text] [doi: 10.3389/fpubh.2020.575091] [Medline: 33102424]
11. Smith L, Jacob L, Butler L, Schuch F, Barnett Y, Grabovac I, et al. Prevalence and correlates of physical activity in a sample of UK adults observing social distancing during the COVID-19 pandemic. BMJ Open Sport Exerc Med 2020 Jul 01;6(1):e000850. [doi: 10.1136/bmjsem-2020-000850]
12. FitBit Staff. The Impact Of Coronavirus On Global Activity. 2020 Mar 23. URL: https://blog.fitbit.com/covid-19-global-activity/ [accessed 2020-12-23]
13. FitBit Staff. Activity bounces back as lockdowns lift, but some healthy gains are slipping. 2020. URL: https://blog.fitbit.com/lockdowns-lift-mobility-changes/ [accessed 2020-12-23]
14. Garmin. Are we still moving? The effect of the global pandemic on active lifestyles. 2020. URL: https://www.garmin.com/en-US/blog/general/the-effect-of-the-global-pandemic-on-active-lifestyles/ [accessed 2020-12-23]
15. Tison GH, Avram R, Kuhar P, Abreau S, Marcus GM, Pletcher MJ, et al. Worldwide Effect of COVID-19 on Physical Activity: A Descriptive Study. Ann Intern Med 2020 Nov 03;173(9):767-770. [doi: 10.7326/m20-2665]
16. BetterPoints. URL: www.BetterPoints.ltd [accessed 2020-12-23]
17. English indices of deprivation 2019. URL: imd-by-postcode.opendatacommunities.org/imd/2019 [accessed 2020-12-23]
18. Hekler EB, Buman MP, Grieco L, Rosenberger M, Winter SJ, Haskell W, et al. Validation of Physical Activity Tracking via Android Smartphones Compared to ActiGraph Accelerometer: Laboratory-Based and Free-Living Validation Studies. JMIR mHealth uHealth 2015 Apr 15;3(2):e36 [FREE Full text] [doi: 10.2196/mhealth.3505] [Medline: 25881662]

19. Duncan MJ, Wunderlich K, Zhao Y, Faulkner G. Walk this way: validity evidence of iphone health application step count in laboratory and free-living conditions. J Sports Sci 2018 Aug;36(15):1695-1704. [doi: 10.1080/02640414.2017.1409855] [Medline: 29179653]
20. Sport England. Active lives adult survey: May 17/18 report. 2018. URL: https://sportengland-production-files.s3.eu-west-2.amazonaws.com/s3fs-public/active-lives-adult-may-17-18-report.pdf [accessed 2020-12-23]
21. Carter EM, Potts HWW. Predicting length of stay from an electronic patient record system: a primary total knee replacement example. BMC Med Inform Decis Mak 2014 Apr 04;14:26 [FREE Full text] [doi: 10.1186/1472-6947-14-26] [Medline: 24708853]
22. Westgate PM, Burchett WW. A Comparison of Correlation Structure Selection Penalties for Generalized Estimating Equations. Am Stat 2017;71(4):344-353 [FREE Full text] [doi: 10.1080/00031305.2016.1200490] [Medline: 30918414]
23. Jia R, Ayling K, Chalder T, Massey A, Broadbent E, Coupland C, et al. Mental health in the UK during the COVID-19 pandemic: cross-sectional analyses from a community cohort study. BMJ Open 2020 Sep 15;10(9):e040620 [FREE Full text] [doi: 10.1136/bmjopen-2020-040620] [Medline: 32933965]
24. Smith L, Jacob L, Butler L, Schuch F, Barnett Y, Grabovac I, et al. Prevalence and correlates of physical activity in a sample of UK adults observing social distancing during the COVID-19 pandemic. BMJ Open Sport Exerc Med 2020 Jul 01;6(1):e000850. [doi: 10.1136/bmjsem-2020-000850]
25. Smith LEAR, Potts HWW, Fear NT, Michie S, Rubin GJ. Worry, behaviour and stigma following UK Government communications during the COVID-19 outbreak: results from three UK surveys. Report to the Department of Health and Social Care. 2020. URL: https://www.gov.uk/government/organisations/department-of-health-and-social-care [accessed 2020-12-23]
26. Ofcom. Communications Market Report 2019. 2019 Jul 04. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0028/155278/communications-market-report-2019.pdf [accessed 2020-10-08]

Abbreviations

GEE: generalized estimating equation
GLM: generalized linear model
PA: physical activity

Edited by G Eysenbach; submitted 21.08.20; peer-reviewed by A Olsen, C Forbes; comments to author 27.09.20; revised version received 09.10.20; accepted 25.10.20; published 03.02.21.

Please cite as:
McCarthy H, Potts HWW, Fisher A
Physical Activity Behavior Before, During, and After COVID-19 Restrictions: Longitudinal Smartphone-Tracking Study of Adults in the United Kingdom
J Med Internet Res 2021;23(2):e23701
URL: https://www.jmir.org/2021/2/e23701
doi:10.2196/23701
PMID:33347421

Original Paper

SARS-CoV-2 Wave Two Surveillance in East Asia and the Pacific: Longitudinal Trend Analysis

Lori Ann Post[1], PhD; Jasmine S Lin[2], BA; Charles B Moss[3], PhD; Robert Leo Murphy[4], MD; Michael G Ison[5], MD, MSc; Chad J Achenbach[5], MD, MPH; Danielle Resnick[6], PhD; Lauren Nadya Singh[1], MPH; Janine White[1], MA; Michael J Boctor[2], BSc; Sarah B Welch[1], MPH; James Francis Oehmke[1], PhD

[1]Buehler Center for Health Policy and Economics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Feinburg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Institute of Food and Agricultural Sciences, University of Florida, Gainsville, FL, United States

[4]Institute for Global Health, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[5]Division of Infectious Disease, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[6]International Food Policy Research Institute, Washington DC, DC, United States

Corresponding Author:
Lori Ann Post, PhD
Buehler Center for Health Policy and Economics
Feinberg School of Medicine
Northwestern University
420 E Superior St
Chicago, IL, 60611
United States
Phone: 1 2039807108
Email: lori.post@northwestern.edu

Abstract

Background: The COVID-19 pandemic has had a profound global impact on governments, health care systems, economies, and populations around the world. Within the East Asia and Pacific region, some countries have mitigated the spread of the novel coronavirus effectively and largely avoided severe negative consequences, while others still struggle with containment. As the second wave reaches East Asia and the Pacific, it becomes more evident that additional SARS-CoV-2 surveillance is needed to track recent shifts, rates of increase, and persistence associated with the pandemic.

Objective: The goal of this study is to provide advanced surveillance metrics for COVID-19 transmission that account for speed, acceleration, jerk, persistence, and weekly shifts, to better understand country risk for explosive growth and those countries who are managing the pandemic successfully. Existing surveillance coupled with our dynamic metrics of transmission will inform health policy to control the COVID-19 pandemic until an effective vaccine is developed. We provide novel indicators to measure disease transmission.

Methods: Using a longitudinal trend analysis study design, we extracted 330 days of COVID-19 data from public health registries. We used an empirical difference equation to measure the daily number of cases in East Asia and the Pacific as a function of the prior number of cases, the level of testing, and weekly shift variables based on a dynamic panel model that was estimated using the generalized method of moments approach by implementing the Arellano-Bond estimator in R.

Results: The standard surveillance metrics for Indonesia, the Philippines, and Myanmar were concerning as they had the largest new caseloads at 4301, 2588, and 1387, respectively. When looking at the acceleration of new COVID-19 infections, we found that French Polynesia, Malaysia, and the Philippines had rates at 3.17, 0.22, and 0.06 per 100,000. These three countries also ranked highest in terms of jerk at 15.45, 0.10, and 0.04, respectively.

Conclusions: Two of the most populous countries in East Asia and the Pacific, Indonesia and the Philippines, have alarming surveillance metrics. These two countries rank highest in new infections in the region. The highest rates of speed, acceleration, and positive upwards jerk belong to French Polynesia, Malaysia, and the Philippines, and may result in explosive growth. While all countries in East Asia and the Pacific need to be cautious about reopening their countries since outbreaks are likely to occur in the second wave of COVID-19, the country of greatest concern is the Philippines. Based on standard and enhanced surveillance, the Philippines has not gained control of the COVID-19 epidemic, which is particularly troubling because the country ranks 4th

in population in the region. Without extreme and rigid social distancing, quarantines, hygiene, and masking to reverse trends, the Philippines will remain on the global top 5 list of worst COVID-19 outbreaks resulting in high morbidity and mortality. The second wave will only exacerbate existing conditions and increase COVID-19 transmissions.

(J Med Internet Res 2021;23(2):e25454) doi:10.2196/25454

KEYWORDS

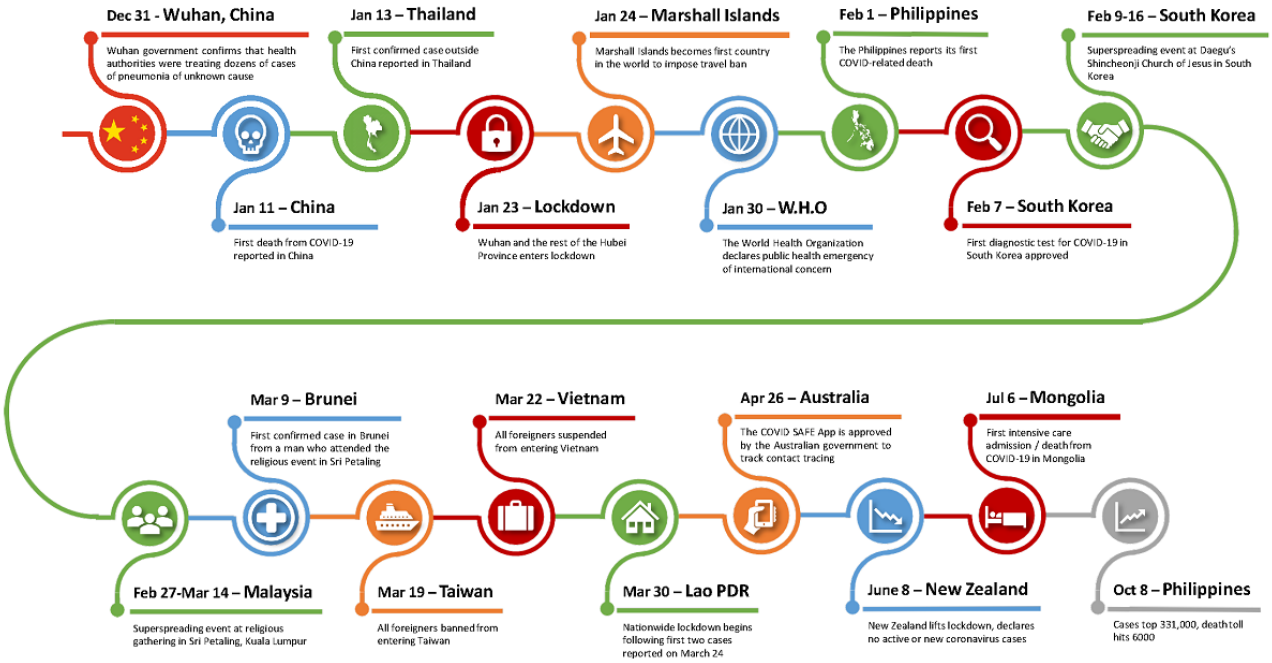
COVID-19; SARS-CoV-2; SARS-CoV-2 surveillance; second wave; wave two; wave 2; global COVID-19 surveillance; Asia Pacific public health surveillance; Asia Pacific COVID-19; Asian Pacific SARS-CoV-2; Asia Pacific surveillance metrics; dynamic panel data; generalized method of the moments; Asian Pacific econometrics; East Asian Pacific COVID-19 surveillance system; Pacific Asian COVID-19 transmission speed; Asian Pacific COVID-19 transmission acceleration; COVID-19 transmission deceleration; COVID-19 transmission jerk; COVID-19 7-day lag; Arellano-Bond estimator; generalized method of moments; GMM; Australia; Brunei; Cambodia; China; Fiji; French Polynesia; Guam; Indonesia; Japan; Kiribati; Laos; Malaysia; Mongolia; Myanmar; New Caledonia; Philippines

Introduction

Background

COVID-19, caused by SARS-CoV-2, was first identified in Wuhan, Hubei Province, China, in December 2019 [1]. Since then, it has spread around the globe including to every country in the East Asia and Pacific region, severely straining governments, health care systems, economies, and quality of life globally (Figure 1). East Asia and the Pacific, as defined by the World Bank, consists of American Samoa, Australia, Brunei Darussalam, Cambodia, China, Fiji, French Polynesia, Guam, Hong Kong, Indonesia, Japan, Kiribati, People's Democratic Republic of Korea, Republic of Korea, Lao People's Democratic Republic, Macao, Malaysia, Marshall Islands, Federated States of Micronesia, Mongolia, Myanmar, Nauru, New Caledonia, New Zealand, Northern Mariana Islands, Palau, Papua New Guinea, the Philippines, Samoa, Singapore, Solomon Islands, Thailand, Timor-Leste, Tonga, Tuvalu, Vanuatu, and Vietnam [2]. Not all of these countries collect or report COVID-19 caseloads and deaths, such as North Korea. As of October 28, 2020, the World Health Organization (WHO) reports a total of 43,540,739 cases and 1,160,650 deaths in these countries [3]. This global region encompasses countries of diverse income levels, political systems, cultures, populations, geography, climate, and health care systems, factors which have profoundly influenced not only the effects of the virus but also the response of each member country.

Figure 1. COVID-19 timeline in East Asia and the Pacific. WHO: World Health Organization. PDR: People's Democratic Republic.



Government Response and Public Health Policy

The extreme quarantine implemented in Wuhan and the entire Hubei region of China mandated all residents to shelter in place without exception [4]. Wuhan imposed travel restrictions in and out of the city in addition to canceling gatherings, closing public places, and shutting down schools and universities [5]. China allocated significant resources for public health service and epidemic prevention and control [6].

The prompt response by East Asian countries was informed by the 2003 severe acute respiratory syndrome (SARS) and 2015 Middle East respiratory syndrome (MERS) outbreaks [7,8]. Singapore and Vietnam's strategy of comprehensive surveillance

to detect and contain as many cases as possible has been highly successful in controlling or eliminating SARS-CoV-2 [9,10]. Taiwan's response to the 2003 SARS epidemic also informed COVID-19 response [11,12]. The Taiwan Centers for Disease Control aggressively traced confirmed cases while the government distributed masks and personal protection equipment [13,14]. Taiwan's interconnected public health, medical, and insurance infrastructure reduces barriers to doctor appointments and follow-up visits, allowing their health care system to capture more cases. Furthermore, their single-payer model allows for centralized health records of population-level longitudinal data, a valuable tool for analyzing the spread of the pandemic [11].

South Korea was the hardest hit country outside of the Middle East during the MERS outbreak in 2015, prompting the Korea Centers for Disease Control (KCDC) to prepare for the next infectious disease outbreak [15]. When COVID-19 breached Korea's borders, the KCDC actively performed contact tracing, quarantined exposed individuals, and diagnosed and isolated new cases with rapid and extensive testing [16-19].

New Zealand eradicated COVID-19 by introducing some of the strictest lockdown measures early on, allowing the government to pursue an elimination approach rather than the typical, mitigation-based model of pandemic planning. Schools and nonessential workplaces were closed, social gatherings banned, and severe travel restrictions applied [20-22]. Australia implemented similar though somewhat less stringent lockdown measures and border closures, resulting in substantially lower crude case fatality and hospitalization rates than many other high-income countries [23,24].

Some low- to middle-income countries such as Vietnam and Mongolia have also been able to implement successful responses to the pandemic despite limited resources or shared borders with China. As a result of Mongolia's travel restrictions, lockdown measures, and surveillance of active infections, they reported no confirmed cases until March 10, 2020, and no intensive care admissions or deaths until July 6, 2020 [25]. Vietnam responded immediately to its first cases by activating an emergency prevention system involving intense surveillance, quarantine, and contact tracing. Those who broke social distancing measures were severely punished. As of July 8, 2020, the nation had the highest test per confirmed case ratio in the world [26-31]. This approach differs vastly from the widespread testing strategy employed by South Korea, which, although seen by many as a best practice in fighting the pandemic, is more resource intensive [32].

In contrast, the Philippines has become Southeast Asia's coronavirus hotspot, overtaking Indonesia with a caseload exceeding 360,000 people with nearly 7000 deaths and no coherent strategy for defeating the virus [33]. Although their first case was reported in January 2020, a national lockdown was not enacted until March, and when citizens took to the streets to protest a lack of food and supplies 2 weeks later, President Rodrigo Duterte threatened that the police and military would shoot those who did not comply with stay-at-home orders [34-36]. Once the lockdown was lifted in June, cases quickly began to climb again. In addition, the Philippines was late to initiate COVID-19 measures including testing, isolation, and contact tracing [33]. Their result paralleled the United States with a similar outcome. The first wave is still raging through the islands while the second wave commences around the globe [37].

The Pacific Islands have been some of the least affected nations in the world due to their unique ability to shut down border traffic [38]. As of August 2, 2020, only 6 countries of the Pacific Islands (Papua New Guinea, Fiji, French Polynesia, Guam, New Caledonia, and Northern Mariana Islands) have recorded positive COVID-19 cases. Among the 12 countries without any confirmed cases, 10 are in this region (Kiribati, Marshall Islands, Micronesia, Nauru, Palau, Samoa, Solomon Islands, Tonga, Tuvalu, and Vanuatu). Furthermore, many nations recently strengthened their infectious disease prevention, surveillance, and response systems due to the re-emergence of measles in the area in 2019 [38-40].

Health Systems, Vulnerable Populations, and Health Disparities

The East Asia and Pacific region encompass a wide range of health care systems with varying capacities for pandemic preparedness. In New Zealand, approximately one-fifth of the government's spending goes to the health sector, and health services are either free or heavily subsidized [41]. In contrast, some of the world's smallest, least developed, and most isolated nations in need of health system strengthening are in the Pacific Islands region [42].

In the Philippines, 43% of the urban population lives in slums, and people living in densely populated urban slums are unlikely to have the space or economic means to practice social distancing [43,44]. The majority of Filipinos pay out of pocket for health care, a prohibitive cost that disproportionately affects the country's 7.5 million senior citizens, many of whom live in rural areas [45]. These barriers to access are exacerbated by the fact that hospitals in the Philippines are currently overwhelmed and reaching maximum capacity [45].

In Cambodia, more than 3 million people lack access to safe water, and 6 million lack access to improved sanitation. This disproportionately affects rural communities, where approximately 77% of Cambodians live, placing them at greater risk from the pandemic [46,47].

Meanwhile, the pandemic has added an additional layer of complexity to Indonesia's longstanding problem of food insecurity and reliance on food imports. In 2018, 55% of the Indonesian population experienced moderate or severe food insecurity. Unemployment and other loss of income associated with the pandemic have likely exacerbated this problem, with 70% of low-income households reporting a loss of income and about the same proportion reporting shortages of some foods or not eating as much as they should, placing 24 million children at risk of food insecurity [48].

Myanmar has attempted to combat the aggravating effects of food insecurity on its already underfunded health sector by providing emergency rations through strategies such as community-based food banks [49,50], although 10%-15% of

the population report consuming reduced quantities of nutritious foods [51].

Individuals with underlying medical conditions are also known to be at greater risk of COVID-19. In Australia, Indigenous Australians constitute a uniquely vulnerable population due to the increased prevalence of diabetes and respiratory and cardiovascular conditions, as well as high reported smoking rates [52,53]. China is notorious for having the worst air pollution problem in the world, which may correlate with susceptibility to respiratory infections [54-56].

Despite early exposure, a high population density, an aging population, and little social distancing measures, Japan reports low rates of infection and death from COVID-19 [57,58]. This has led to some hypotheses that the Bacille Calmette-Guérin (BCG) vaccine against tuberculosis may protect against the virus, as countries that mandate the BCG vaccine have relatively low per capita death rates from COVID-19 [57,59].

Economy

Tourism is an important source of revenue for many economies in developing Asia and the Pacific [50,60,61]. For countries like Palau, where international tourism receipts are close to 50% of the GDP (gross domestic product) and over a third of international tourists are from China, the decline in tourism due to COVID-19 has been devastating [60].

China is a major trade destination for many developing Asian economies such as Mongolia, the Philippines, Singapore, Taipei, and Vietnam [60]. Long quarantine-like conditions have the potential to deeply harm export-based economies such as Mongolia, where coal exports were reduced due to border restrictions [25].

Culture

While Western cultures endorse individualism and a more independent self-concept, Eastern cultures emphasize collectivism and a concept of the self as interdependent with others, which may motivate individuals to remain committed to COVID-19 precautions even at the expense of personal freedoms [62]. Some of the actions that have enabled Asian countries to contain the spread of the virus have been challenging to Western notions of privacy and individual freedom; such measures have been almost universally accepted in Asian countries [26,32]. Therefore, interpersonal transmission of the virus may be less likely in East Asian countries [63].

Public health departments, as well as universities and media outlets, are tracking the novel coronavirus using raw data, including the number of new infections, testing, positivity, R_0 (reproduction number), deaths, local hospital capacity, etc [39,64-93]. Public health surveillance informs policy on "flattening the curve" of COVID-19 [94-97]. Epidemiologists have utilized various modeling techniques to forecast the numbers of cases and deaths attributed to the virus [98-102]. Both the WHO and the Center for Systems Science and Engineering at Johns Hopkins University have developed tracking tools [98]. While helpful, these static metrics suffer from incomplete case ascertainment and data contamination [94,96]. Existing surveillance is a proxy for the true coronavirus caseload because public health surveillance systems tend to pick up the most severe cases [103,104], which is especially problematic when tracking SARS-CoV-2 because most carriers are asymptomatic, presymptomatic, or only have mild symptoms [105-108]. Public health surveillance that can control for these limitations are needed. Moreover, metrics that detect how transmission speed of the novel coronavirus, shifts in the pandemic, acceleration in speed, and persistence of COVID-19 based on prior infections are needed to supplement existing measures [83].

Objective

The objective of our research is to use a longitudinal trend analysis study design in concert with dynamic panel modeling and method of moments to correct for existing surveillance data limitations [94,96]. Specifically, we will measure significant weekly shifts in the increase, decrease, or plateaued transmission of SARS-CoV-2. Our study will measure the underlying causal effect from last week that persists through this week, with a 7-day persistence rate to explain a clustering/declustering effect. The 7-day persistence rate represents an underlying disease transmission wave, where a large number of transmissions that resulted in a large number of infections today then "echoes" forward into a large number of new transmissions and hence a large number of new cases 7 days later. If positive, it is consistent with, for example, a mega-event (eg, the large religious gathering at Daegu's Shincheonji Church of Jesus in South Korea) that causes an increase in the number of cases in adjoining days, among other explanations [16]. If zero or nonsignificant, it is indicative of a constant rate of new infections and/or a constant size of the infectious population. In summary, we will measure negative and positive shifts in the transmission of SARS-CoV-2 or acceleration/deceleration rates that are not limited by sampling bias.

Methods

We conducted a longitudinal trend analysis for our study design using data extracted from the internet. The COVID Tracking Project [109], Our World in Data [110], and The Foundation for Innovative New Diagnostics [111] compiles data from multiple sources on the web [112]; data for the most recent 4 weeks were accessed from the GitHub repository [113-115]. This resulted in a panel of 26 countries in East Asia and the Pacific with 30 days in each panel (n=780). An empirical difference equation was specified in which the number of positive cases in each country for each day is a function of the prior number of cases, the level of testing, and weekly shift variables that measure whether the contagion was growing faster, at the same rate, or slower than the previous weeks. This resulted in a dynamic panel model that was estimated using the generalized method of moments (GMM) approach by implementing the Arellano-Bond estimator in R (The R Foundation) [94,96,116]. Additionally, we report on the novel dynamic surveillance metrics of speed, acceleration, and jerk [94,96].

Results

Country Regression Results

Regression results are presented for 26 East Asian and Pacific countries in Table 1. Weekly surveillance data in Tables 2-6 and Figure 2 [117] are based on these regressions.

The Wald statistic for the regression was significant (χ^2_5=49,836,424; *P*<.001). The Sargan test was not significant, failing to reject the validity of overidentifying restrictions (χ^2_{294}=18; *P*=.99).

Table 1. Arellano-Bond dynamic panel data modeling of the number of daily infections reported by country in East Asia and the Pacific, October 5-18, 2020.

Variable	Statistic	*P* value
L1Pos[a]	*r*=–0.007	.89
L7Pos[b]	*r*=0.887	<.001
Cumulative tests	*r*=0.000	.15
Weekend	*r*=–0.603	<.001
Wald statistic for regression	χ^2_5=49836424	<.001
Sargan statistic for validity	χ^2_{294}=18	.99

[a]L1Pos: the statistical impact of a 1-day lag of speed on today's value of speed.

[b]L7Pos: the statistical impact of the 7-day lag of speed on today's value of speed. New cases per day tend to have an echo effect 7 days later. Reported as the weekly average number of new cases per day that are attributable to the weekly average of the 7-day lag of the number of new cases per day.

Figure 2. COVID-19 weekly trends in East Asia and the Pacific [117].

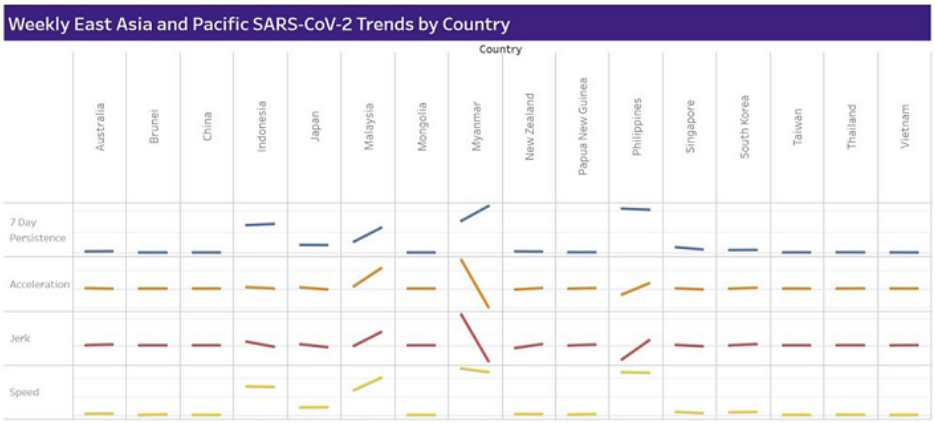


Table 2, for the week of October 5-11, and Table 3, for the week of October 12-18, present traditional surveillance metrics including new COVID-19 cases, cumulative COVID-19 cases, 7-day moving average of COVID-19 infections, infections per 100,000 population, deaths, cumulative deaths, and 7-day moving average of deaths rates per 100,000 population. Overall, in East Asia and the Pacific, by the second week, there were 9951 new daily cases of COVID-19, 1,076,042 cumulative cases of COVID-19, a 7-day moving average of 8005, an infection rate per 100,000 population of 0.4344, 216 daily deaths, and 28,053 cumulative deaths.

For the week of October 5-11, Indonesia led the East Asian and Pacific region with the highest number of new cases at 4294, followed by the Philippines at 2156 and Myanmar at 2158. The number of deaths follow the same pattern for Indonesia, the Philippines, and Myanmar at 88, 86, and 32 deaths, respectively.

For the week of October 12-18 (Table 3), Indonesia, the Philippines, and Myanmar continued to lead the region with 4301, 2588, and 1287 new COVID-19 infections and 84, 72, and 39 deaths, respectively. The cumulative number of deaths for the region reached 28,053 by the second week in October.

Table 2. Static surveillance metrics for the week of October 5-11, 2020.

Country	New COVID-19 cases, n	Cumulative COVID-19 cases, n	7-day moving average of new cases	Rate of infection	New deaths, n	Cumulative deaths, n	Death rate per 100k
Australia	15	27,244	17.57	0.06	0	897	0
Brunei	0	146	0	0	0	3	0
China	21	90,778	24.86	0	0	4739	0
French Polynesia	0	2692	104	0	0	10	0
Guam	89	3078	54.14	53.20	2	60	1.20
Indonesia	4294	328,952	4207	1.59	88	11,765	0.03
Japan	679	88,912	510.43	0.54	3	1627	0
Malaysia	374	15,096	429.71	1.17	3	155	0.01
Mongolia	0	318	0	0	3	0	0.09
Myanmar	2158	26,064	1365.86	3.99	32	598	0.06
New Zealand	1	1515	2.43	0.02	0	25	0
Papua New Guinea	0	549	1.29	0	0	7	0
Philippines	2156	336,926	2514	1.99	86	6238	0.08
Singapore	7	1,010,010	9.43	0.12	0	27	0
South Korea	58	57,866	73.57	0.11	2	432	0
Taiwan	0	24,606	1.43	0	0	7	0
Thailand	0	527	7.29	0	0	59	0
Vietnam	2	3634	1.57	0	0	35	0
Region	9854	1107	7999	0.43	219	26,684	0.01

Table 3. Static surveillance metrics for the week of October 12-18, 2020.

Country	New COVID-19 cases, n	Cumulative COVID-19 cases, n	7-day moving average of new cases	Rate of infection	New deaths, n	Cumulative deaths, n	Death rate per 100k
Australia	12	27,383	19.86	0.05	0	904	0
Brunei	0	147	0.14	0	0	3	0
China	30	90,955	25.29	0	0	4739	0
French Polynesia	62	3797	157.9	22.20	0	14	0
Guam	0	3617	77.00	0	3	66	1.79
Indonesia	4301	357,762	4115.71	1.59	84	12,431	0.03
Japan	593	92,656	534.9	0.47	9	1670	0.01
Malaysia	869	19,627	647.3	2.72	4	180	0.01
Mongolia	0	324	0	0	4	0	0.12
Myanmar	1387	34,875	1258.71	2.57	39	838	0.07
New Zealand	3	1530	2.14	0.06	0	25	0
Papua New Guinea	3	581	4.57	0.03	0	7	0
Philippines	2588	354,338	2487.4	2.39	72	6603	0.07
Singapore	3	57,904	5.43	0.05	0	28	0
South Korea	91	25,199	84.71	0.18	1	444	0
Taiwan	0	535	1.14	0	0	7	0
Thailand	7	3686	7.43	0.01	0	59	0
Vietnam	2	1126	2.71	0	0	35	0
Region	9951	1,076,042	8005	0.44	216	28,053	0.01

Tables 4 and 5 provide the novel surveillance metrics for the weeks of October 5-11 and October 12-18, respectively. During the week of October 5-11 (Table 4), French Polynesia had the highest speed or velocity of new cases at 37 per 100,000 population, followed by Guam at 32.6 cases per 100,000 population. The highest rates of acceleration per 100,000 population was 0.312 for Myanmar, 0.025 for Malaysia, and 0.015 for Indonesia. The highest jerk rates were 0.228, 0.027, and 0.008 for Myanmar, Indonesia, and Japan, respectively. French Polynesia and Guam ranked 1 and 2 for 7-day persistence at 28.68 and 25.14, respectively, meaning these cases were statically attributed to those persons infected 7 days earlier.

Table 4. Novel surveillance metrics for the week of October 5-11, 2020.

Country	Speed[a]	Acceleration[b]	Jerk[c]	7-day persistence effect on speed[d]
Australia	0.07	0	0	0.05
Brunei	0	0	0	0
China	0	0	0	0
French Polynesia	37.24	0	–11.66	28.69
Guam	32.36	–5.12	–9.82	26.14
Indonesia	1.55	0.02	0.03	1.32
Japan	0.40	0.01	0.01	0.37
Malaysia	1.34	0.03	0	0.52
Mongolia	0	0	0	0
Myanmar	2.53	0.31	0.23	1.53
New Zealand	0.05	–0.01	–0.02	0.05
Papua New Guinea	0.01	0	0	0.01
Philippines	2.33	–0.07	–0.11	2.12
Singapore	0.17	0	0	0.26
South Korea	0.14	0	0	0.12
Taiwan	0.01	0	0	0
Thailand	0.01	0	0	0.01
Vietnam	0	0	0	0
Region	0.41	0.01	0	0.33

[a]Daily positives per 100k (weekly average of new daily cases per 100k).

[b]Day-to-day change in the number of positives per day, weekly average, per 100k.

[c]Week-over-week change in acceleration, per 100k.

[d]New cases per day per 100k attributed to new cases 7 days ago.

Table 5 presents the novel surveillance metrics for the second week of our study period. Between October 12 and 18, French Polynesia ranked first in speed of new infections at 56.5 per 100,000, followed by Guam at 46 per 100,000. French Polynesia and Guam are several standard deviations higher than the rest of the East Asian and Pacific region. French Polynesia had the highest acceleration rate at 3.17 per 100,000 population, and Polynesia had the highest positive jerk at 15.4 per 100,000 population. French Polynesia and Guam had the largest 7-day persistence during the October 12-18 period (Table 6). In summary, French Polynesia, Malaysia, the Philippines, South Korea, and New Zealand have positive speeds, acceleration, jerks, and 7-day persistence, indicating an upwards shift in the pandemic.

Table 5. Novel surveillance metrics for the week of October 12-18, 2020.

Country	Speed[a]	Acceleration[b]	Jerk[c]	7-day persistence effect on speed[d]
Australia	0.08	0	0.01	0.06
Brunei	0.03	0	0	0
China	0	0	0	0
French Polynesia	56.52	3.17	15.45	33.04
Guam	46.03	–7.60	–19.13	28.72
Indonesia	1.52	0	–0.01	1.38
Japan	0.42	–0.01	–0.01	0.36
Malaysia	2.03	0.22	0.10	1.19
Mongolia	0	0	0	0
Myanmar	2.33	–0.20	–0.12	2.24
New Zealand	0.04	0.01	0.01	0.04
Papua New Guinea	0.05	0	0	0.01
Philippines	2.30	0.06	0.04	2.06
Singapore	0.10	–0.01	–0.01	0.15
South Korea	0.16	0.01	0.01	0.13
Taiwan	0	0	0	0.01
Thailand	0.01	0	0	0.01
Vietnam	0	0	0	0
Region	0.41	0	0	0.36

[a]Daily positives per 100k (weekly average of new daily cases per 100k).

[b]Day-to-day change in the number of positives per day, weekly average, per 100k.

[c]Week-over-week change in acceleration, per 100k.

[d]New cases per day per 100k attributed to new cases 7 days ago.

Table 6. Seven-day persistence difference.

Country	7-day persistence	
	October 11, 2020	October 18, 2020
French Polynesia	28.69	33.04
Guam	26.14	28.72
Philippines	2.12	2.06
Myanmar	1.53	2.24
Indonesia	1.32	1.38

The most populous countries in East Asia and Pacific include China, Indonesia, Japan, Philippines, and Vietnam (Table 7). Countries with larger populations are at risk for having more COVID-19 infections by virtue of size, but this was not necessarily the case when comparing population size to the speed, acceleration, jerks, and 7-day persistence in Tables 4 and 5.

For comprehensive surveillance of static or traditional surveillance metrics with novel surveillance metrics for East Asia and Pacific, see Multimedia Appendices 1-3.

Table 7. Most populous East Asian countries.

Country[a]	Population as of 2020, N
China	1,439,323,776
Indonesia	273,523,615
Japan	126,476,461
Philippines	109,581,078
Vietnam	97,338,579

[a]Does not include countries that do not track or report COVID-19 cases (eg, North Korea).

Discussion

Countries in the East Asia and Pacific region have had differential success in combating the COVID-19 pandemic, with some countries among the most successful in the world at containing the pandemic and others in serious jeopardy. China, Taiwan, South Korea, and Japan had early outbreaks that were successfully contained through stringent protective measures. Some of the smaller islands in the Pacific encountered the disease much later than other countries, but after initial exposure, French Polynesia and Malaysia had outbreaks that swiftly affected their nations.

While South Korea, New Zealand, and Australia are demonstrating some increases in speed, acceleration, jerk, and 7-day persistence, these nations had successfully implemented COVID-19 control policies that literally eliminated COVID-19. However, as wave two of the COVID-19 pandemic has recently began the cycle of transmissions, these countries had either low caseloads or zero caseloads; hence, any new cases are going to result in increased rates of transmission. The new cases in countries that previously eradicated COVID-19 indicate a need to reinstate public health guidelines to keep the second wave of COVID-19 transmission from gaining a larger foothold. Moreover, other countries in East Asia and the Pacific who are presently dealing with the first wave will have novel infections from wave two as well, such as Indonesia and the Philippines who have the highest cumulative infections and cumulative deaths from COVID-19. While the Philippines was decelerating during the week of October 5-11, it reversed course and now has a positive acceleration.

The region's biggest successes are Singapore, New Zealand, Vietnam, South Korea, and China. In theory, it is likely easier to control for outbreaks in island nations such as Singapore and New Zealand; however, not all islands experienced their level of prevention and mitigation. Moreover, relative to population size (Table 7), South Korea, Vietnam, and China experienced an initial COVID-19 outbreak and took preventative measures to control further spread successfully.

Acknowledgments

This publication was made possible through support provided by Feed the Future through the US Agency for International Development, under the terms of contract #7200LA1800003. The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of the US Agency for International Development.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Weekly East Asia and Pacific 7-day persistence map.
[PNG File , 331 KB - jmir_v23i2e25454_app1.png]

Multimedia Appendix 2
Weekly East Asia and Pacific acceleration jerk map.
[PNG File , 392 KB - jmir_v23i2e25454_app2.png]

Multimedia Appendix 3
Weekly East Asia and Pacific statistics.
[PNG File , 265 KB - jmir_v23i2e25454_app3.png]

References

1. COVID-19 Overview and Infection Prevention and Control Priorities in Non-US Healthcare Settings. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/non-us-settings/overview/index.html [accessed 2020-08-06]
2. East Asia and Pacific. The World Bank. URL: https://data.worldbank.org/region/east-asia-and-pacific [accessed 2020-08-02]
3. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. URL: https://covid19.who.int/table [accessed 2020-10-28]
4. Cyranoski D. What China's coronavirus response can teach the rest of the world. Nature 2020 Mar 17;579(7800):479-480. [doi: 10.1038/d41586-020-00741-x] [Medline: 32203360]
5. Lau H, Khosrawipour V, Kocbach P, Mikolajczyk A, Schubert J, Bania J, et al. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. J Travel Med 2020 May 18;27(3):37-51 [FREE Full text] [doi: 10.1093/jtm/taaa037] [Medline: 32181488]
6. Zhang S, Wang Z, Chang R, Wang H, Xu C, Yu X, et al. COVID-19 containment: China provides important lessons for global response. Front Med 2020 Apr 25;14(2):215-219 [FREE Full text] [doi: 10.1007/s11684-020-0766-9] [Medline: 32212059]
7. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020 Apr 2;5(4):536-544 [FREE Full text] [doi: 10.1038/s41564-020-0695-z] [Medline: 32123347]
8. Frieden TR, Lee CT. Identifying and Interrupting Superspreading Events-Implications for Control of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg Infect Dis 2020 Jun;26(6):1059-1066 [FREE Full text] [doi: 10.3201/eid2606.200495] [Medline: 32187007]
9. Ng Y, Li Z, Chua YX, Chaw WL, Zhao Z, Er B, et al. Evaluation of the Effectiveness of Surveillance and Containment Measures for the First 100 Patients with COVID-19 in Singapore - January 2-February 29, 2020. MMWR Morb Mortal Wkly Rep 2020 Mar 20;69(11):307-311 [FREE Full text] [doi: 10.15585/mmwr.mm6911e1] [Medline: 32191691]
10. Lee VJ, Chiew CJ, Khong WX. Interrupting transmission of COVID-19: lessons from containment efforts in Singapore. J Travel Med 2020 May 18;27(3):39 [FREE Full text] [doi: 10.1093/jtm/taaa039] [Medline: 32167146]
11. Lin C, Braund WE, Auerbach J, Chou J, Teng J, Tu P, et al. Policy Decisions and Use of Information Technology to Fight COVID-19, Taiwan. Emerg Infect Dis 2020 Jul;26(7):1506-1512 [FREE Full text] [doi: 10.3201/eid2607.200574] [Medline: 32228808]
12. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. JAMA 2020 Apr 14;323(14):1341-1342. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
13. Steinbrook R. Contact Tracing, Testing, and Control of COVID-19-Learning From Taiwan. JAMA Intern Med 2020 Sep 01;180(9):1163-1164. [doi: 10.1001/jamainternmed.2020.2072] [Medline: 32356871]
14. Chang CM, Tan TW, Ho TC, Chen CC, Su TH, Lin CY. COVID-19: Taiwan's epidemiological characteristics and public and hospital responses. PeerJ 2020;8:e9360 [FREE Full text] [doi: 10.7717/peerj.9360] [Medline: 32551205]
15. Her M. Repurposing and reshaping of hospitals during the COVID-19 outbreak in South Korea. One Health 2020 Dec;10:100137 [FREE Full text] [doi: 10.1016/j.onehlt.2020.100137] [Medline: 32373705]
16. Choi JY. COVID-19 in South Korea. Postgrad Med J 2020 Jul 04;96(1137):399-402. [doi: 10.1136/postgradmedj-2020-137738] [Medline: 32366457]
17. Kim KH, Choi EH, Kim SK. Editorial. COVID-19 outbreak and its countermeasures in the Republic of Korea. J Neurosurg 2020 Apr 17:1-2 [FREE Full text] [doi: 10.3171/2020.4.JNS201062] [Medline: 32302997]
18. Her M. How is COVID-19 affecting South Korea? What is our current strategy? Disaster Med Public Health Prep 2020 Apr 03:1-7. [doi: 10.1017/dmp.2020.69] [Medline: 32241325]
19. Choi S, Han C, Lee J, Kim S, Kim IB. Innovative screening tests for COVID-19 in South Korea. Clin Exp Emerg Med 2020 Jun 16;7(2):73-77 [FREE Full text] [doi: 10.15441/ceem.20.032] [Medline: 32295319]
20. Cousins S. New Zealand eliminates COVID-19. The Lancet 2020 May;395(10235):1474. [doi: 10.1016/s0140-6736(20)31097-7]
21. Klein A. New Zealand's success. New Scientist 2020 May;246(3282):9. [doi: 10.1016/s0262-4079(20)30908-8]
22. Varghese C, Xu W. Quantifying what could have been - The impact of the Australian and New Zealand governments' response to COVID-19. Infect Dis Health 2020 Nov;25(4):242-244 [FREE Full text] [doi: 10.1016/j.idh.2020.05.003] [Medline: 32507662]
23. Andrikopoulos S, Johnson G. The Australian response to the COVID-19 pandemic and diabetes - Lessons learned. Diabetes Res Clin Pract 2020 Jul;165:108246 [FREE Full text] [doi: 10.1016/j.diabres.2020.108246] [Medline: 32502693]
24. Antony J, James WT, Neriamparambil AJ, Barot DD, Withers T. An Australian Response to the COVID-19 Pandemic and Its Implications on the Practice of Neurosurgery. World Neurosurg 2020 Jul;139:e864-e871 [FREE Full text] [doi: 10.1016/j.wneu.2020.05.136] [Medline: 32450310]
25. Erkhembayar R, Dickinson E, Badarch D, Narula I, Warburton D, Thomas GN, et al. Early policy actions and emergency response to the COVID-19 pandemic in Mongolia: experiences and challenges. The Lancet Global Health 2020 Sep;8(9):e1234-e1241. [doi: 10.1016/s2214-109x(20)30295-3]

26. Dao TL, Nguyen TD, Hoang VT. Controlling the COVID-19 pandemic: Useful lessons from Vietnam. Travel Med Infect Dis 2020 Sep;37:101822 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101822] [Medline: 32653477]
27. Duong DM, Le VT, Ha BTT. Controlling the COVID-19 Pandemic in Vietnam: Lessons From a Limited Resource Country. Asia Pac J Public Health 2020 May 19;32(4):161-162. [doi: 10.1177/1010539520927290] [Medline: 32429676]
28. Ha BTT, Ngoc Quang L, Mirzoev T, Tai NT, Thai PQ, Dinh PC. Combating the COVID-19 Epidemic: Experiences from Vietnam. Int J Environ Res Public Health 2020 Apr 30;17(9):3125 [FREE Full text] [doi: 10.3390/ijerph17093125] [Medline: 32365851]
29. Hoang VM, Hoang HH, Khuong QL, La NQ, Tran TTH. Describing the pattern of the COVID-19 epidemic in Vietnam. Glob Health Action 2020 Dec 31;13(1):1776526 [FREE Full text] [doi: 10.1080/16549716.2020.1776526] [Medline: 32588779]
30. Nguyen TH, Vu DC. Summary of the COVID-19 outbreak in Vietnam - Lessons and suggestions. Travel Med Infect Dis 2020 Sep;37:101651 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101651] [Medline: 32247928]
31. Van Nguyen H, Van Hoang M, Dao ATM, Nguyen HL, Van Nguyen T, Nguyen PT, et al. An adaptive model of health system organization and responses helped Vietnam to successfully halt the Covid-19 pandemic: What lessons can be learned from a resource-constrained country. Int J Health Plann Manage 2020 Sep 18;35(5):988-992 [FREE Full text] [doi: 10.1002/hpm.3004] [Medline: 32557815]
32. Trevisan M, Le LC, Le AV. The COVID-19 Pandemic: A View From Vietnam. Am J Public Health 2020 Aug;110(8):1152-1153. [doi: 10.2105/ajph.2020.305751]
33. Sullivan M. The Philippines Becomes Coronavirus Hot Spot In Southeast Asia. NPR. 2020 Aug 10. URL: https://www.npr.org/2020/08/10/900279722/the-philippines-becomes-coronavirus-hot-spot-in-southeast-asia [accessed 2020-08-13]
34. Lo B, Ott H. The Philippines has the worst coronavirus outbreak in Southeast Asia, and critics warn the government could use it to crack down on dissent. CBS News. 2020 Aug 07. URL: https://www.cbsnews.com/news/covid-19-philippines-worst-coronavirus-outbreak-in-southeast-asia-crackdown-on-dissent/ [accessed 2020-08-13]
35. Westcott B, Lagamayo A. Critics are accusing the Philippines government of using the coronavirus lockdown to crack down on dissent. CNN. 2020 Aug 03. URL: https://www.cnn.com/2020/08/03/asia/philippines-coronavirus-duterte-intl-hnk/index.html [accessed 2020-08-13]
36. Capatides C. “Shoot them dead”: Philippine President Rodrigo Duterte orders police and military to kill citizens who defy coronavirus lockdown. CBS News. 2020 Apr 02. URL: https://www.cbsnews.com/news/rodrigo-duterte-philippines-president-coronavirus-lockdown-shoot-people-dead/ [accessed 2020-08-13]
37. Moghnieh R, Abdallah D, Bizri AR. COVID-19: Second Wave or Multiple Peaks, Natural Herd Immunity or Vaccine - We Should be Prepared. Disaster Med Public Health Prep 2020 Sep 10:1-8 [FREE Full text] [doi: 10.1017/dmp.2020.349] [Medline: 32907693]
38. Diarra I, Muna L, Diarra U. How the Islands of the South Pacific have remained relatively unscathed in the midst of the COVID-19 pandemic. J Microbiol Immunol Infect 2020 Jul 10:6-15 [FREE Full text] [doi: 10.1016/j.jmii.2020.06.015] [Medline: 32665107]
39. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Center for Systems Science and Engineering at Johns Hopkins University. 2020. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-09-11]
40. Kerbaj C, Cazorla C, De Greslan T, Serie M, Gourinat AC, Marot B. COVID-19: The New Caledonia Experience. Clin Infect Dis 2020 Nov 19;71(16):2279-2281 [FREE Full text] [doi: 10.1093/cid/ciaa600] [Medline: 32415955]
41. Bandyopadhyay G, Meltzer A. Let us unite against COVID-19 - a New Zealand perspective. Ir J Psychol Med 2020 May 14:1-4 [FREE Full text] [doi: 10.1017/ipm.2020.44] [Medline: 32404225]
42. Craig AT, Heywood AE, Hall J. Risk of COVID-19 importation to the Pacific islands through global air travel. Epidemiol Infect 2020 Mar 23;148:e71 [FREE Full text] [doi: 10.1017/S0950268820000710] [Medline: 32202489]
43. Lau LL, Hung N, Go DJ, Ferma J, Choi M, Dodd W, et al. Knowledge, attitudes and practices of COVID-19 among income-poor households in the Philippines: A cross-sectional study. J Glob Health 2020 Jun;10(1):011007 [FREE Full text] [doi: 10.7189/jogh.10.011007] [Medline: 32566169]
44. World Development Indicators. World Bank Group. 2020. URL: https://databank.worldbank.org/source/world-development-indicators [accessed 2020-08-13]
45. Buenaventura RD, Ho JB, Lapid MI. COVID-19 and mental health of older adults in the Philippines: a perspective from a developing country. Int Psychogeriatr 2020 Oct;32(10):1129-1133 [FREE Full text] [doi: 10.1017/S1041610220000757] [Medline: 32349826]
46. Cambodia's water and sanitation crisis. Water.org. 2020. URL: https://water.org/our-impact/where-we-work/cambodia/ [accessed 2020-08-02]
47. Water, Sanitation and Hygiene. UNICEF Cambodia. 2020. URL: https://www.unicef.org/cambodia/water-sanitation-and-hygiene [accessed 2020-08-02]
48. COVID-19 Economic and Food Security Implications for Indonesia. WFP. 2020. URL: https://www.wfp.org/publications/covid-19-economic-and-food-security-implications-indonesia-3rd-edition-august-2020 [accessed 2020-11-01]

49. COVID-19 Policy Response (CPR) Portal. International Food Policy Research Institute. 2020. URL: https://www.ifpri.org/project/covid-19-policy-response-cpr-portal [accessed 2020-08-14]
50. Grundy-Warr C, Lin S. COVID-19 geopolitics: silence and erasure in Cambodia and Myanmar in times of pandemic. Eurasian Geography and Economics 2020 Jun 16;61(4-5):493-510. [doi: 10.1080/15387216.2020.1780928]
51. Impacts of COVID-19 on Myanmar's agri-food system: Evidence base and policy implications. International Food Policy Research Institute. 2020. URL: https://www.ifpri.org/publication/impacts-covid-19-myanmars-agri-food-system-evidence-base-and-policy-implications [accessed 2020-08-03]
52. Desborough J, Hall Dykgraaf S, de Toca L, Davis S, Roberts L, Kelaher C, et al. Australia's national COVID-19 primary care response. Med J Aust 2020 Aug 04;213(3):104-106.e1 [FREE Full text] [doi: 10.5694/mja2.50693] [Medline: 32623740]
53. Furlong Y, Finnie T. Culture counts: the diverse effects of culture and society on mental health amidst COVID-19 outbreak in Australia. Ir J Psychol Med 2020 Sep 14;37(3):237-242 [FREE Full text] [doi: 10.1017/ipm.2020.37] [Medline: 32406358]
54. Xu H, Yan C, Fu Q, Xiao K, Yu Y, Han D, et al. Possible environmental effects on the spread of COVID-19 in China. Sci Total Environ 2020 Aug 20;731:139211 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.139211] [Medline: 32402910]
55. Kan H, Chen B, Hong C. Health impact of outdoor air pollution in China: current knowledge and future research needs. Environ Health Perspect 2009 May;117(5):A187 [FREE Full text] [doi: 10.1289/ehp.12737] [Medline: 19478975]
56. Zhu Y, Xie J, Huang F, Cao L. Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. Science of The Total Environment 2020 Jul 20;727:138704. [doi: 10.1016/j.scitotenv.2020.138704]
57. Iwasaki A, Grubaugh ND. Why does Japan have so few cases of COVID-19? EMBO Mol Med 2020 May 08;12(5):e12481 [FREE Full text] [doi: 10.15252/emmm.202012481] [Medline: 32275804]
58. Hamidi S, Sabouri S, Ewing R. Does Density Aggravate the COVID-19 Pandemic? Journal of the American Planning Association 2020 Jun 18;86(4):495-509. [doi: 10.1080/01944363.2020.1777891]
59. O'Neill LAJ, Netea MG. BCG-induced trained immunity: can it offer protection against COVID-19? Nat Rev Immunol 2020 Jun 11;20(6):335-337 [FREE Full text] [doi: 10.1038/s41577-020-0337-y] [Medline: 32393823]
60. Abiad A, Arao RM, Dagil S, Ferrarini B, Noy I, Osewe P, et al. The Economic Impact of the COVID-19 Outbreak on Developing Asia. Asian Development Bank 2020 Mar:1-14 [FREE Full text] [doi: 10.22617/BRF200096]
61. Yamano T, Pradhananga M, Schipani S, Samson JNG, Quiao L, Leuangkhamsing S, et al. The Impact of COVID-19 on Tourism Enterprises in the Lao People's Democratic Republic: An Initial Assessment. Asian Development Bank 2020 Jun:1-8 [FREE Full text] [doi: 10.22617/brf200187-2]
62. Burgess A, Horii M. Risk, ritual and health responsibilisation: Japan's 'safety blanket' of surgical face mask-wearing. Sociol Health Illn 2012 Nov;34(8):1184-1198 [FREE Full text] [doi: 10.1111/j.1467-9566.2012.01466.x] [Medline: 22443378]
63. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May 30;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
64. Coronavirus World Map: Tracking the Global Outbreak. The New York Times. 2020. URL: https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html [accessed 2020-08-12]
65. Tracking coronavirus' global spread. CNN. 2020. URL: https://www.cnn.com/interactive/2020/health/coronavirus-maps-and-cases/ [accessed 2020-08-12]
66. Coronavirus Disease-19, Republic of Korea. Ministry of Health and Welfare of Korea. 2020. URL: http://ncov.mohw.go.kr/en/ [accessed 2020-08-12]
67. COVID-19 (SARS-CoV-2-infection). Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov.tw/En [accessed 2020-08-12]
68. COVID-19 - current cases. Ministry of Health of New Zealand. 2020. URL: https://www.health.govt.nz/our-work/diseases-and-conditions/covid-19-novel-coronavirus/covid-19-current-situation/covid-19-current-cases [accessed 2020-08-12]
69. About Coronavirus Disease 2019 (COVID-19). Ministry of Health, Labour, and Welfare – Japan. 2020. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/newpage_00032.html [accessed 2020-08-12]
70. Coronavirus (COVID-19) health alert. Australian Government Department of Health. 2020. URL: https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert [accessed 2020-08-14]
71. COVID-19 Dashboard. Brunei Darussalam Ministry of Health. 2020. URL: www.healthinfo.gov.bn [accessed 2020-08-14]
72. COVID-19 Information. US Embassy in Cambodia. 2020. URL: https://kh.usembassy.gov/covid-19-information/ [accessed 2020-08-14]
73. Aug 14: Daily briefing on novel coronavirus cases in China. National Health Commission of the People's Republic of China. 2020. URL: http://en.nhc.gov.cn/2020-08/14/c_81388.htm [accessed 2020-08-14]
74. COVID-19 Updates. Fiji Ministry of Health and Medical Services. 2020. URL: http://www.health.gov.fj/covid-19-updates/ [accessed 2020-08-14]
75. Guam COVID-19 Dashboard. Department of Public Health and Social Services. 2020. URL: http://dphss.guam.gov/covid-19/ [accessed 2020-08-14]
76. Coronavirus Disease (COVID-19) in Hong Kong. The Government of the Hong Kong Special Administrative Region. 2020. URL: https://www.coronavirus.gov.hk/eng/index.html [accessed 2020-08-14]

77. COVID-19. Ministry of Health Republic of Indonesia. 2020. URL: https://www.kemkes.go.id/index.php?lg=LN02 [accessed 2020-08-14]
78. Ad hoc committee COVID-19. Ministry of Health – Laos. 2020. URL: https://www.covid19.gov.la/index.php [accessed 2020-08-14]
79. Home | COVID-19 Malaysia. Ministry of Health of Malaysia. 2020. URL: http://covid-19.moh.gov.my/ [accessed 2020-08-14]
80. RMI Covid-19 Update. Marshall Islands National Disaster Management Office. 2020. URL: https://ndmo.gov.mh/ [accessed 2020-08-14]
81. COVID-19 Information. US Embassy in the Federated States of Micronesia. 2020. URL: https://fm.usembassy.gov/covid-19-information/ [accessed 2020-08-14]
82. COVID-19. Ministry of Health of Mongolia. 2020. URL: https://covid19.mohs.mn/ [accessed 2020-08-14]
83. COVID-19 (Coronavirus Disease 2019) Acute Respiratory Disease. The Republic of the Union of Myanmar- Ministry of Health and Sports. 2020. URL: https://www.mohs.gov.mm/Main/content/publication/2019-ncov [accessed 2020-08-14]
84. COVID-19. Department of Health and Medicinal Service. 2020. URL: http://www.naurugov.nr/government/departments/department-of-health-and-medicinal-service/covid-19.aspx [accessed 2020-08-14]
85. CNMI COVID-19 Dashboard. Commonwealth Healthcare Corporation – Commonwealth of the Northern Mariana Islands. 2020. URL: https://cnmichcc.maps.arcgis.com/apps/opsdashboard/index.html#/4061b674fc964efe84f7774b7979d2b5 [accessed 2020-08-14]
86. COVID-19. Ministry of Health Republic of Palau. 2020. URL: http://www.palauhealth.org/ [accessed 2020-08-14]
87. COVID-19 Pandemic. Ministry of Health – Papua New Guinea. 2020. URL: https://www.health.gov.pg/subindex.php?news=1 [accessed 2020-08-14]
88. COVID-19 Tracker. Republic of the Philippines Department of Health. 2020. URL: https://www.doh.gov.ph/covid19tracker [accessed 2020-08-14]
89. Novel Coronavirus (COVID-19). Government of Samoa. 2020. URL: https://www.samoagovt.ws/category/novel-coronavirus-covid-19/ [accessed 2020-08-14]
90. Learn about Coronavirus. Solomon Islands Government. 2020. URL: https://solomons.gov.sb/ministry-of-health-medical-services/essential-services/learn-about-coronavirus/ [accessed 2020-08-14]
91. Updates on COVID-19 (Coronavirus Disease 2019) Local Situation. Ministry of Health of Singapore. 2020. URL: https://www.moh.gov.sg/covid-19 [accessed 2020-08-14]
92. Coronavirus Disease (COVID-19). Thailand Department of Disease Control. 2020. URL: https://ddc.moph.go.th/viralpneumonia/eng/index.php [accessed 2020-08-14]
93. COVID-19 Updates. Vanuatu Ministry of Health. 2020. URL: https://covid19.gov.vu/ [accessed 2020-08-14]
94. Oehmke J, Moss C, Singh L, Oehmke T, Post L. Dynamic Panel Surveillance of COVID-19 Transmission in the United States to Inform Health Policy: Observational Statistical Study. J Med Internet Res 2020 Oct 5;22(10):e21955 [FREE Full text] [doi: 10.2196/21955] [Medline: 32924962]
95. Ibrahim NK. Epidemiologic surveillance for controlling Covid-19 pandemic: types, challenges and implications. J Infect Public Health 2020 Nov;13(11):1630-1638 [FREE Full text] [doi: 10.1016/j.jiph.2020.07.019] [Medline: 32855090]
96. Oehmke JF, Oehmke TB, Singh LN, Post LA. Dynamic Panel Estimate-Based Health Surveillance of SARS-CoV-2 Infection Rates to Inform Public Health Policy: Model Development and Validation. J Med Internet Res 2020 Sep 22;22(9):e20924 [FREE Full text] [doi: 10.2196/20924] [Medline: 32915762]
97. Foddai A, Lubroth J, Ellis-Iversen J. Base protocol for real time active random surveillance of coronavirus disease (COVID-19) - Adapting veterinary methodology to public health. One Health 2020 Jun;9:100129 [FREE Full text] [doi: 10.1016/j.onehlt.2020.100129] [Medline: 32292815]
98. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases 2020 May;20(5):533-534. [doi: 10.1016/s1473-3099(20)30120-1]
99. Lin Y, Liu C, Chiu Y. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun 2020 Jul;87:30-32 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.020] [Medline: 32283286]
100. Sajadi MM, Habibzadeh P, Vintzileos A, Shokouhi S, Miralles-Wilhelm F, Amoroso A. Temperature, Humidity and Latitude Analysis to Predict Potential Spread and Seasonality for COVID-19. SSRN Journal Preprint posted online Mar 9, 2020. [doi: 10.2139/ssrn.3550308] [Medline: 32714105]
101. Binti Hamzah FA, Lau C, Nazri H, Ligot DV, Lee G, Tan GL. CoronaTracker: Worldwide COVID-19 Outbreak Data Analysis and Prediction. Bulletin of the World Health Organization 2020 Mar 19:1-32. [doi: 10.2471/BLT.20.255695]
102. Petropoulos F, Makridakis S. Forecasting the novel coronavirus COVID-19. PLoS One 2020 Mar 31;15(3):e0231236 [FREE Full text] [doi: 10.1371/journal.pone.0231236] [Medline: 32231392]
103. Thacker SB, Berkelman RL. Public health surveillance in the United States. Epidemiol Rev 1988;10:164-190. [doi: 10.1093/oxfordjournals.epirev.a036021] [Medline: 3066626]
104. Lee L, Teutsch S, Tacker S, St. Louis M. Considerations in planning a surveillance system. In: Principles and Practice of Public Health Surveillance. Oxford, UK: Oxford Scholarship Online; 2010:18-28.

105. Day M. Covid-19: four fifths of cases are asymptomatic, China figures indicate. BMJ 2020 Apr 02;369:m1375. [doi: 10.1136/bmj.m1375] [Medline: 32241884]
106. He J, Guo Y, Mao R, Zhang J. Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis. J Med Virol 2021 Feb 13;93(2):820-830 [FREE Full text] [doi: 10.1002/jmv.26326] [Medline: 32691881]
107. Al-Sadeq DW, Nasrallah GK. The incidence of the novel coronavirus SARS-CoV-2 among asymptomatic patients: A systematic review. Int J Infect Dis 2020 Sep;98:372-380 [FREE Full text] [doi: 10.1016/j.ijid.2020.06.098] [Medline: 32623083]
108. An P, Song P, Wang Y, Liu B. Asymptomatic Patients with Novel Coronavirus Disease (COVID-19). Balkan Med J 2020 Jun 01;37(4):229-230 [FREE Full text] [doi: 10.4274/balkanmedj.galenos.2020.2020.4.20] [Medline: 32279479]
109. Lipton Z, Riley K. The Covid Tracking Project. The Atlantic. 2020. URL: https://covidtracking.com/ [accessed 2020-10-19]
110. Roser M, Ortiz-Ospina E, Hassell J. Coronavirus Pandemic (COVID-19). Our World in Data. 2020. URL: https://ourworldindata.org/coronavirus [accessed 2020-10-13]
111. FINDCov19TrackerData. The Foundation for Innovative New Diagnostics. 2020. URL: https://www.finddx.org/covid-19/test-tracker/ [accessed 2020-10-13]
112. Brownstein JS, Freifeld CC, Madoff LC. Digital Disease Detection — Harnessing the Web for Public Health Surveillance. N Engl J Med 2009 May 21;360(21):2153-2157. [doi: 10.1056/nejmp0900702]
113. The Covid Tracking Project. GitHub. 2020. URL: https://github.com/COVID19Tracking [accessed 2020-10-19]
114. FINDCov19TrackerData. GitHub. 2020. URL: https://github.com/dsbbfinddx/FINDCov19TrackerData [accessed 2020-10-19]
115. COVID-19 Data Repository by their for Systems Science and Engineering (CSSE) at Johns Hopkins University. GitHub. 2020. URL: https://github.com/CSSEGISandData/COVID-19 [accessed 2020-10-19]
116. Hansen LP. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 1982 Jul;50(4):1029. [doi: 10.2307/1912775]
117. Post L. Global SARS-CoV-2 Surveillance:Policy, Persistence, Transmission-East Asia and Pacific. 2020 Oct 25. URL: https://public.tableau.com/profile/lori.post#!/vizhome/GlobalSARS-CoV-2SurveillancePolicyPersistenceandTransmission-EastAsiaandPacific/Story [accessed 2020-10-25]

Abbreviations

BCG: Bacille Calmette-Guérin
GDP: gross domestic product
GMM: generalized method of moments
KCDC: Korean Centers for Disease Control
MERS: Middle East respiratory syndrome
SARS: severe acute respiratory syndrome
WHO: World Health Organization

Edited by G Eysenbach; submitted 09.11.20; peer-reviewed by P Banik; comments to author 30.11.20; revised version received 30.11.20; accepted 16.01.21; published 01.02.21.

Please cite as:
Post LA, Lin JS, Moss CB, Murphy RL, Ison MG, Achenbach CJ, Resnick D, Singh LN, White J, Boctor MJ, Welch SB, Oehmke JF
SARS-CoV-2 Wave Two Surveillance in East Asia and the Pacific: Longitudinal Trend Analysis
J Med Internet Res 2021;23(2):e25454
URL: https://www.jmir.org/2021/2/e25454
doi:10.2196/25454
PMID:33464207

Original Paper

Estimated Sleep Duration Before and During the COVID-19 Pandemic in Major Metropolitan Areas on Different Continents: Observational Study of Smartphone App Data

Rebecca Robbins[1,2], PhD; Mahmoud Affouf[3], PhD; Matthew D Weaver[1,2], PhD; Mark É Czeisler[4,5], AB; Laura K Barger[1,2], PhD; Stuart F Quan[1,2], MD; Charles A Czeisler[1,2], MD, PhD, FRCP

[1]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA, United States

[2]Division of Sleep Medicine, Harvard Medical School, Boston, MA, United States

[3]Department of Mathematics, Kean University, Union, NJ, United States

[4]School of Psychological Sciences, Turner Institute Brain and Mental Health, Monash University, Victoria, Australia

[5]Institute for Breathing and Sleep, Austin Health, Melbourne, Australia

Corresponding Author:
Rebecca Robbins, PhD
Division of Sleep and Circadian Disorders
Departments of Medicine and Neurology
Brigham and Women's Hospital
221 Longwood Avenue
Boston, MA, 02115
United States
Phone: 1 2039792338
Email: rrobbins4@bwh.harvard.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/2/e28057/

Abstract

Background: Amid the COVID-19 pandemic, public health policies to curb the spread of SARS-CoV-2 and its associated disease, COVID-19, have resulted in significant alterations to daily routines (eg, work-from-home policies) that may have enabled longer sleep duration among the general population.

Objective: We aimed to examine changes in estimated sleep duration in 5 major metropolitan areas before and after the start of the COVID-19 pandemic.

Methods: We conducted a prospective observational study using estimated sleep duration data obtained from a smartphone app. The data were obtained from regular users of the smartphone app before and after the World Health Organization declared COVID-19 a pandemic in March 2020. We compared within-subject estimated sleep duration before and during the COVID-19 pandemic using generalized linear mixed models.

Results: Among the 2,871,037 observations, 957,022 (33.3%) were from users in London; 549,151 (19.1%) were from users in Los Angeles; 846,527 (29.5%) were from users in New York City; 251,113 (8.7%) were from users in Seoul; and 267,224 (9.3%) were from users in Stockholm. The average age of the users in the sample was 35 years (SE 11 years). Prior to the COVID-19 pandemic, people residing in Seoul had the shortest estimated sleep duration (mean 6 hours 28 minutes, SE 11.6 minutes) and those residing in Stockholm had the longest estimated sleep duration (mean 7 hours 34 minutes, SE 9.9 minutes). The onset of the COVID-19 pandemic was associated with a 13.7 minute increase in estimated sleep duration when comparing March 2019 and March 2020 (95% CI 13.1-14.3, $P<.001$) and an increase of 22.3 minutes when comparing April 2019 and April 2020 (95% CI 21.5-23.1, $P<.001$).

Conclusions: The average estimated sleep duration increased sharply in the months after the onset of the COVID-19 pandemic. This finding suggests that the implementation of COVID-19 mitigation strategies has provided people worldwide with increased opportunities to sleep, which may enhance the response of the immune system to viral pathogens.

(J Med Internet Res 2021;23(2):e20546) doi:10.2196/20546

KEYWORDS

sleep health; mobile health; sleep tracking; COVID-19; sleep; observational study; app

Introduction

Public health officials worldwide have implemented stringent measures to curb the spread of SARS-CoV-2 and its associated disease, COVID-19. In some regions, actions to mitigate COVID-19 have been drastic, such as mandatory shelter-in-place regulations, while regulations in other regions have been more lenient [1]. In either case, life has changed markedly for much of the global population.

Research conducted amid crises of similar magnitude to the COVID-19 pandemic has shown that sleep is disrupted during and after such events [2-5]. For instance, research conducted after the 2003 severe acute respiratory syndrome (SARS) outbreak in China demonstrated an increase in insomnia symptoms associated with the onset of the outbreak [4]. In the context of a natural disaster, researchers found that 40% of people who survived an earthquake in Japan reported sleep difficulties in the years following the disaster, and 8% reported short sleep duration [6]. Although previous literature would suggest that sleep duration is likely to decline, sleep duration may increase worldwide during the COVID-19 pandemic for several reasons. First, due to the highly contagious nature of SARS-CoV-2 and the lack of a vaccine, social distancing and work-from-home recommendations and policies have been widely implemented to curb the spread of the virus [7]. As much of the global population is spending less time commuting, more time at home, and less time socializing, it is possible that their sleep duration will increase, contrary to what has been observed during previous crises.

Sleep is a critical element of immune system function [8,9]. Experimental studies have shown that inadequate sleep results in increased susceptibility to viral infection [10,11]. Research has also shown a heightened ability to mount an immune response among those who obtain a healthy, sufficient duration of sleep (ie, 7-9 hours) [12-14]. As the COVID-19 pandemic unfolds, surveillance of sleep duration may be important to identify poor sleep practices and to develop evidence-based interventions and campaigns to enhance sleep in response to this crisis as necessary. To further our understanding of sleep during the COVID-19 pandemic, we analyzed smartphone app–estimated sleep durations of individuals residing in London, England; Los Angeles, California, United States; New York City, New York, United States; Seoul, South Korea; and Stockholm, Sweden, before the onset of the COVID-19 pandemic (January 2019 through April 2019) and after the onset of the pandemic (January 2020 to April 2020).

Methods

Participants

We conducted a prospective observational study using data obtained from the smartphone-based sleep tracking software app Sleep Cycle. We obtained data from regular users of the app who tracked sleep on 80% or more of the days between January 1, 2019 and April 12, 2020 (at least 374/468 days). To understand geographic variations in sleep during the COVID-19 pandemic, we obtained data from individuals living in 5 major metropolitan areas (London, Los Angeles, New York City, Seoul, and Stockholm).

The selection of geographical regions was guided by several factors. First, urban regions were included in this study because there is a lower density of users in rural regions, which could have hindered comparison between rural and urban regions. Second, there were interesting differences in COVID-19 prevalence, preparedness, and mitigation strategies by geographic region. For instance, Sweden did not limit social mobility among its residents as strictly as other regions, such as the United States. Additionally, South Korea had experience combating a pandemic from the 2003 SARS outbreak. Therefore, we requested data from large urban centers on 3 different continents with the aim of exploring different patterns in estimated sleep duration from country to country that may be reflective in part of different prevalence, preparedness, or mitigation strategies implemented in these various countries.

The Sleep Cycle app logs a participant's place of residence by the coordinates identified by their smartphone GPS, which participants agree to provide when using the app. Only users who agreed to Sleep Cycle's privacy policy, which dictates that data may be used for research purposes, were included in the data set.

Estimated Sleep Duration

Users open the app when they go to bed and first attempt to sleep; they place their smartphone either next to their bed or on their mattress, and they close the app either after the built-in alarm clock wakes them or when they wake naturally from sleep. Sleep is calculated by the app as the time interval between these two digital sleep diary events (first attempt to sleep and waking) [15], in accordance with the 2016 Consumer Technology Association standards for wearable sleep monitors [16], herein termed "estimated sleep duration." We obtained an anonymized data set that included dated estimated sleep durations for each user between January 1, 2019 and April 12, 2020. The data set comprised 2,974,922 observations. We removed observations with a duration shorter than 1 hour from the data set to mitigate the impact of napping on the estimated sleep duration. These shorter observations represented 3.5% of the sample (103,885/2,974,922 observations). Removal of these shorter observations resulted in a final data set with 2,871,037 observations for analysis.

Statistical Analyses

Descriptive statistics were used to characterize the demographic characteristics of the sample. To evaluate population-level changes in estimated sleep duration during the COVID-19 pandemic while accounting for established seasonal variation,

we compared the estimated sleep duration by geographic location in the same calendar month between years using generalized linear mixed models. This method included hierarchical random effects for the city and study participants, which accounted for the dependence between repeated measures and the clustering of respondents in each location. This design enabled us to compare estimated sleep duration over time before and after the COVID-19 pandemic (eg, March 2019 vs March 2020) in each geographic location. All analyses were performed in Stata Statistical Software for Mac Version 16 (StataCorp LLC).

Results

In total, 2,871,037 nights from 8218 unique users between January 1, 2019 and April 12, 2020 were available for analysis. Of the 2,871,037 estimated sleep durations analyzed, 957,022 (33.3%) were from London; 549,151 (19.1%) were from Los Angeles; 846,527 (29.5%) were from New York City; 251,113 (8.7%) were from Seoul; and 267,224 (9.3%) were from Stockholm. The average age of users in the sample was 35 years (SE 11 years). Prior to the COVID-19 pandemic, those residing in Seoul had the shortest average estimated sleep duration (mean 6 hours 28 minutes, SE 11.6 minutes) and those residing in Stockholm had the longest average estimated sleep duration (mean 7 hours 34 minutes, SE 9.9 minutes); the average estimated sleep durations were 7 hours 32 minutes (SE 5.5 minutes) in London, 7 hours 21 minutes (SE 8.4 minutes) in Los Angeles and 7 hours 22 minutes (SE 6.3 minutes) in New York City. Table 1 displays descriptive demographic and estimated sleep duration characteristics by city prior to the COVID-19 pandemic.

Table 1. Descriptive statistics summarizing the participants' demographic characteristics and sleep data by city prior to the COVID-19 pandemic (N=2,871,037 nights; N=8,218 users).

Characteristic	Value	95% CI	SE
London (n=957,022 nights)			
Age (years), mean	34.4	34.3-34.4	0.0
Gender (n=2687 users), n (%)			
Male	1423 (53.5)	51.8-55.6	0.0
Female	1236 (46.5)	44.4-48.2	0.0
Estimated sleep duration	7 hours 32 minutes	7 hours 32 minutes–7 hours 33 minutes	5.5 minutes
Los Angeles (n=549,151 nights)			
Age (years), mean	36.0	36.1-36.1	0.02
Gender (n=1452 users), n (%)			
Male	787 (54.2)	51.6-56.8	0.0
Female	665 (45.8)	43.2-48.4	0.0
Estimated sleep duration	7 hours 21 minutes	7 hours 21 minutes–7 hours 22 minutes	8.4 minutes
New York City (n=846,527 nights)			
Age (years), mean	35.0	34.9-35.1	0.0
Gender (n=2398 users), n (%)			
Male	1209 (50.6)	48.6-52.6	0.0
Female	1182 (49.4)	47.4-51.4	0.0
Estimated sleep duration	7 hours 22 minutes	7 hours 22 minutes–7 hours 22 minutes	6.3 minutes
Seoul (n=251,113 nights)			
Age (years), mean	33.8	33.8-33.8	0.0
Gender (n=809 users), n (%)			
Male	488 (60.3)	56.9-63.6	0.0
Female	321 (39.7)	36.4-43.1	0.0
Estimated sleep duration	6 hours 28 minutes	6 hours 28 minutes–6 hours 28 minutes	11.6 minutes
Stockholm (n=267,224 nights; n=868 users)			
Age (years)	40.7	40.7-40.8	0.0
Gender			
Male	450 (51.8)	48.5-55.2	0.0
Female	418 (48.2)	44.8-51.5	0.0
Estimated sleep duration	7 hours 34 minutes	7 hours 33 minutes–7 hours 34 minutes	9.9 minutes

As shown in Figures 1 and 2, there was an increase in estimated sleep duration starting in February 2020 compared to February 2019, which coincided with the onset of the COVID-19 pandemic in the 5 cities across 3 continents.

Figure 1. Estimated sleep duration by month (January to April) in 2019 and in 2020 for the total sample. The I-bars represent the 95% CIs.

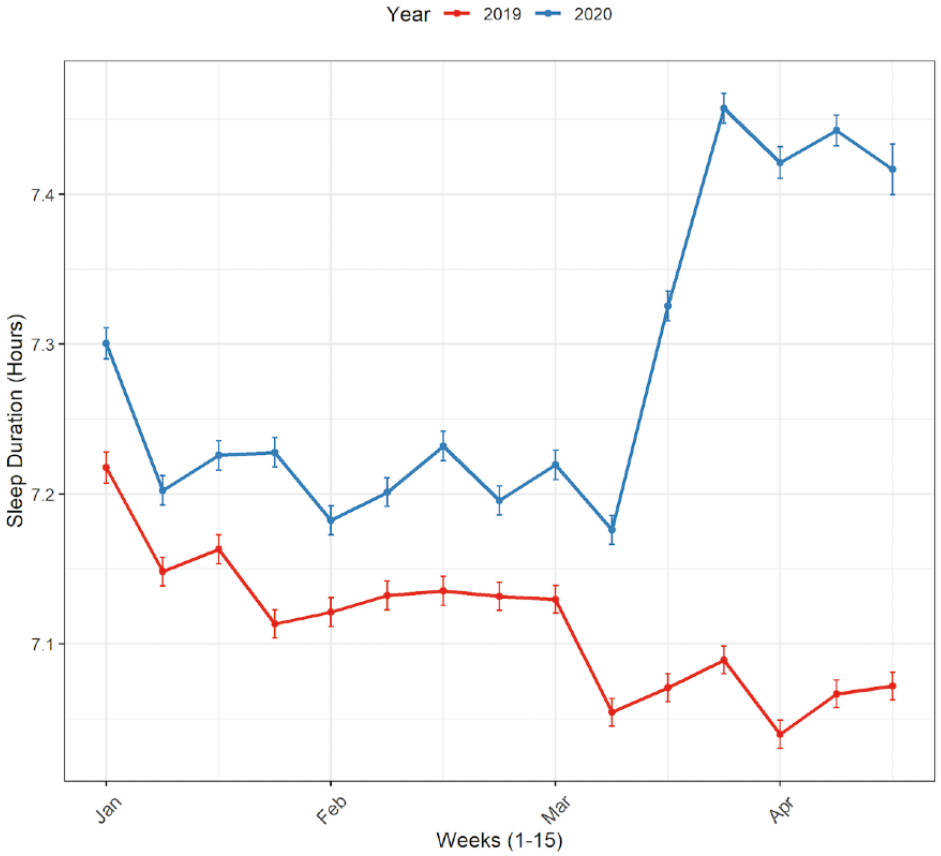
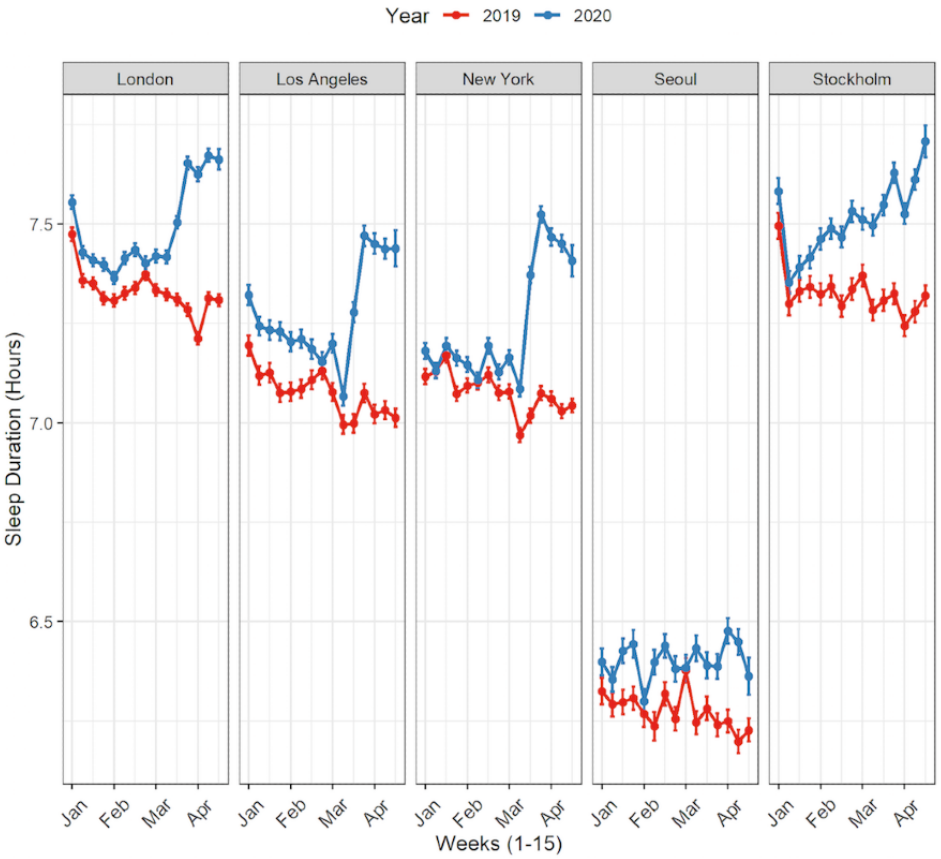


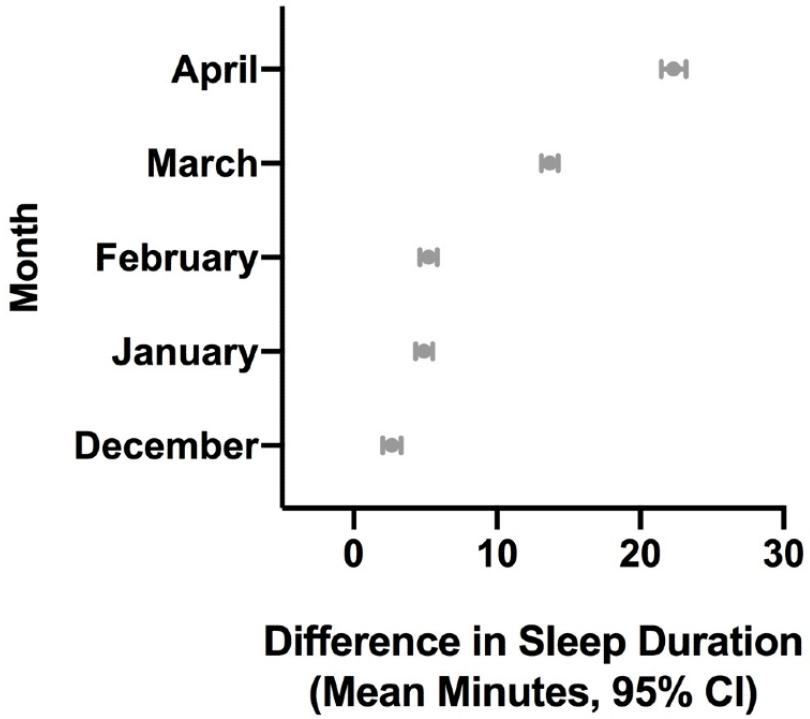
Figure 2. Estimated sleep duration by months (January to April) in 2019 and in 2020 for the total sample by city.



Results from the generalized linear mixed models are displayed in Figures 3-8. As shown in Figure 3, estimated sleep duration increased after the onset of COVID-19 in the total sample. In the total sample, comparing January 2019 to January 2020, we

observed increases of 4.9 minutes of estimated sleep duration (95% CI 4.3-5.5 minutes; *P*<.001); comparing February 2019 to February 2020, we observed increases of 5.2 minutes of estimated sleep duration (95% CI 4.81-5.91 minutes; *P*<.001); comparing March 2019 to March 2020, we observed increases of 13.7 minutes of estimated sleep duration (95% CI 13.1-14.3 minutes; *P*<.001); and when comparing April 2019 to April 2020, we observed increases of 22.3 minutes of estimated sleep duration (95% CI 21.5-23.2 minutes; *P*<.001) (Figure 3).

Figure 3. Difference in estimated sleep duration during the COVID-19 pandemic compared to the same month in the previous year in the total sample.



Models examining estimated sleep duration after the COVID-19 outbreak by city are displayed in Figures 4-8. Seoul had the smallest increase when comparing April 2019 and April 2020 (12.2 minutes, 95% CI 9.5-14.4 minutes; *P*<.001), while New York had the largest (24.5 minutes, 95% CI 23.1-25.9 minutes, *P*<.001), followed by London (20.8 minutes, 95% CI 19.5-22.1 minutes; *P*<.001).

Figure 4. Difference in estimated sleep duration during the COVID-19 pandemic compared to the same month in the previous year in London.

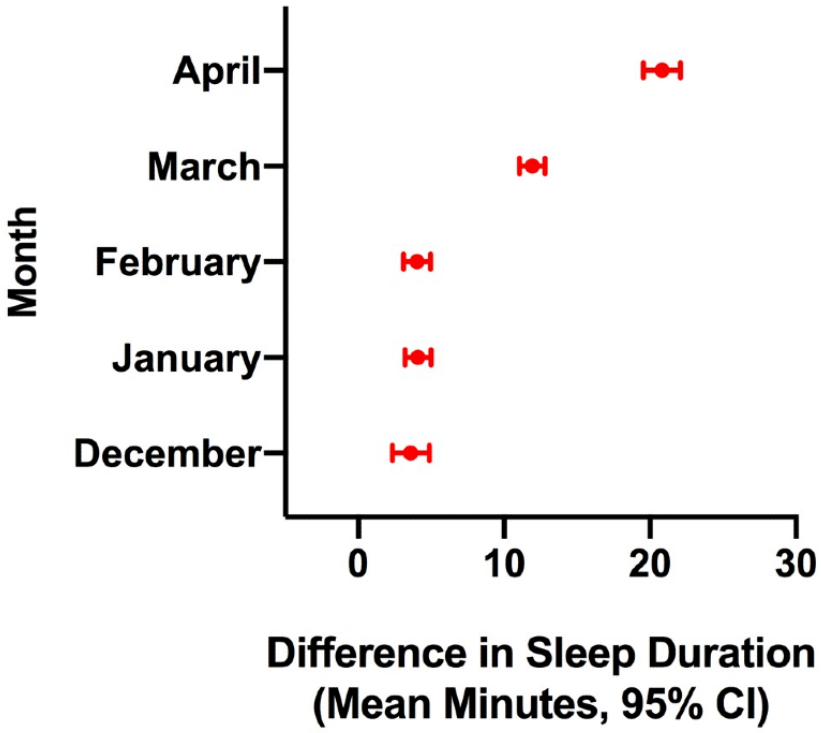


Figure 5. Difference in estimated sleep duration during the COVID-19 pandemic compared to the same month in the previous year in Los Angeles.

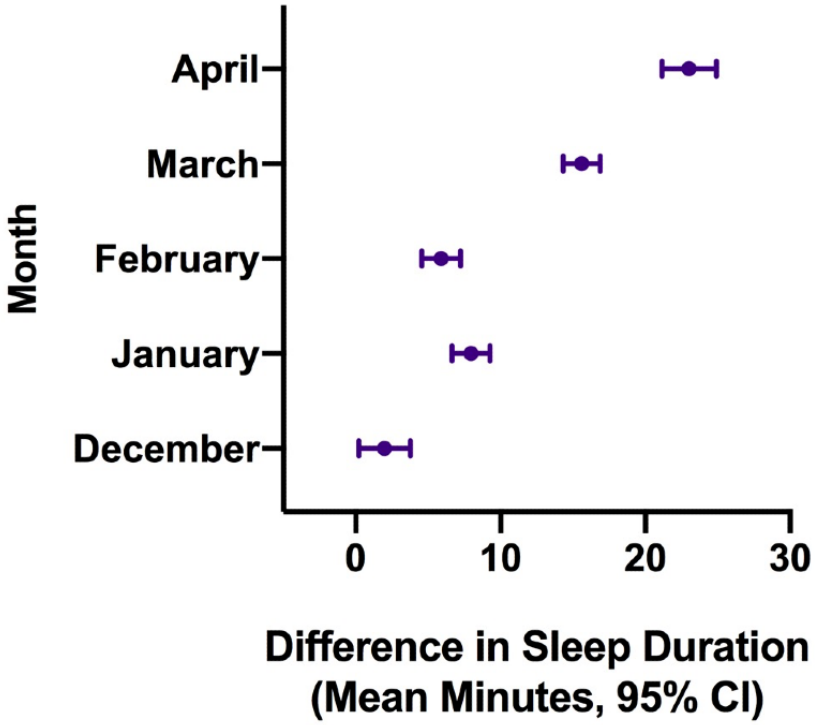


Figure 6. Difference in estimated sleep duration during the COVID-19 pandemic compared to the same month in the previous year in New York.

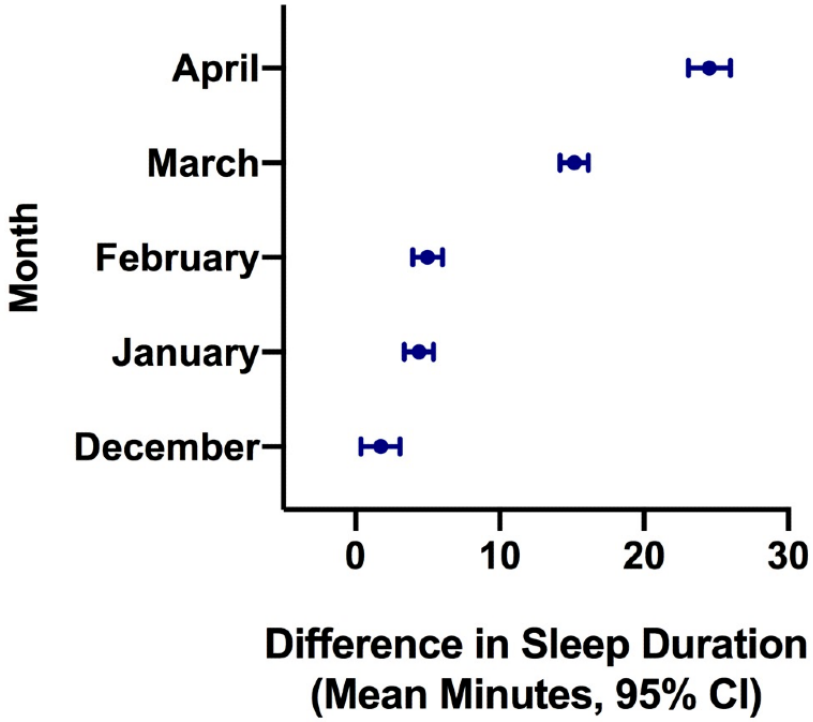


Figure 7. Difference in estimated sleep duration during the COVID-19 pandemic compared to the same month in the previous year in Seoul.

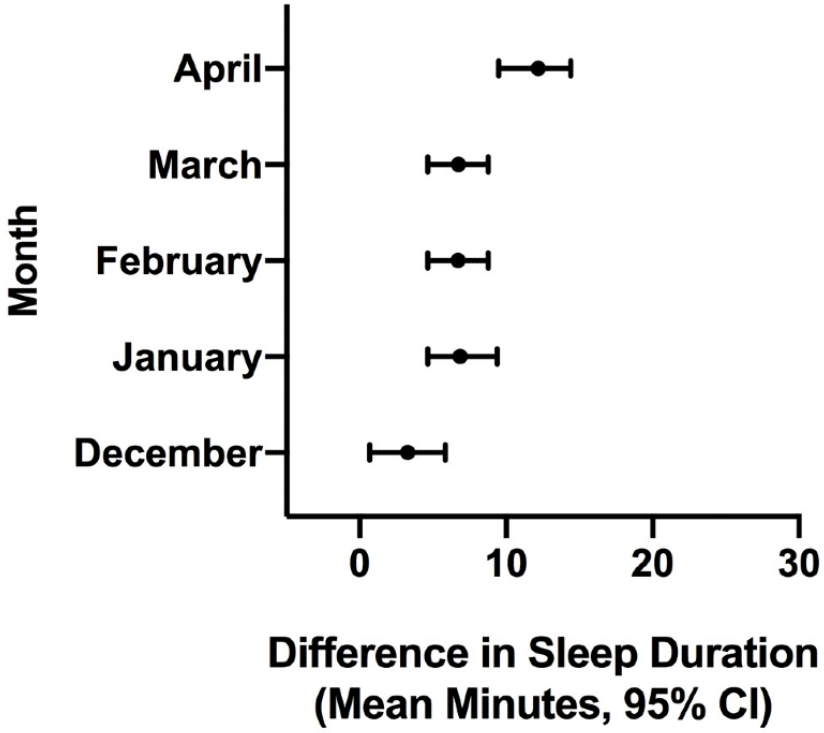
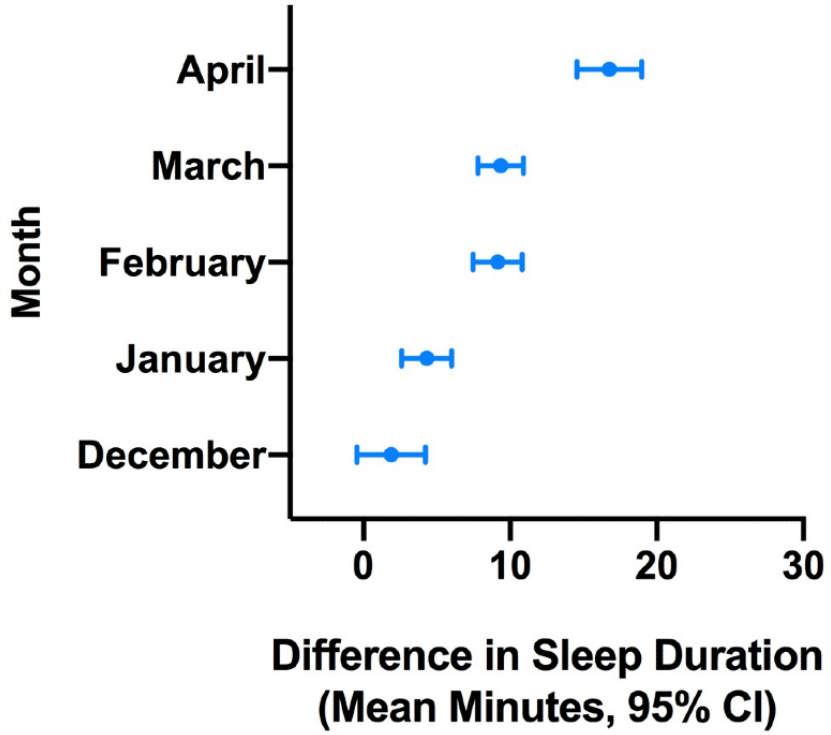


Figure 8. Difference in estimated sleep duration during the COVID-19 pandemic compared to the same month in the previous year in Stockholm.



Discussion

Principal Findings

The results of this study, based on 2.9 million nightly recordings of smartphone-estimated sleep duration, demonstrate an abrupt, significant increase in estimated sleep duration in the months after COVID-19 was declared a pandemic and international health crisis by the World Health Organization [17] compared with the same months from the same individuals in the prior year. Further, the average estimated sleep duration in several regions was at or below the lower range of the recommendation for healthy sleep before the pandemic (ie, 7 hours or less), even among those concerned about sleep enough to track it; however, this duration increased to be within the range of healthy, sufficient sleep amid the onset of the COVID-19 pandemic. Our data, gathered from a daily sleep tracker smartphone app, reveal that on average there were significant increases in estimated sleep duration in London, Los Angeles, New York City, Seoul, and Stockholm concomitant with the onset of the COVID-19 pandemic, consistent with data from subjective, retrospective self-reports among 425 adults in 3 European countries [18]. We hypothesize that these increases in smartphone app–estimated sleep duration, together with the recently reported increases in sleep and wake time regularity following the onset of the COVID-19 pandemic [19], have occurred as a consequence of stay-at-home orders, nonessential business closures, and work-from-home policies implemented to slow community spread of the infection. Sufficient sleep duration is a critical element of immune system functioning, with inadequate sleep resulting in increased susceptibility to viral infection and reduced antibody production after vaccination [8,9]. We cannot exclude the possibility that individuals in this study were experiencing longer times in bed as opposed to longer estimated sleep duration. Nevertheless, if the observed increase in estimated sleep duration, as detected by the app, in fact represents an increase in actual sleep amid the COVID-19 pandemic, this may have been an important factor in reducing the risk of progression to COVID-19 following exposure to SARS-CoV-2 [10,11] and in enhancing the ability of people with COVID-19 to mount an effective immune response [12,13].

Additionally, we detected notable differences in sleep by geographic region. The average estimated sleep duration prior to the COVID-19 pandemic of app users in Seoul was >1 hour shorter than those in London and Stockholm, which is consistent with prior research, and >40 minutes shorter than those in New York City and Los Angeles. This is consistent with prior research, which documented that sleep deprivation is common in South Korea [20,21]. Moreover, app users in Seoul also demonstrated the smallest increase in estimated sleep duration during the COVID-19 pandemic, at 11 minutes per night compared to the increases of 20 minutes in other cities. The smaller increase in Seoul compared to other cities could be related to the outbreak of SARS in South Korea in 2003, which may have afforded an opportunity for the country to prepare and contain the spread of the virus without the more drastic mitigation measures, such as extended stay-at-home regulations, that were levied in other countries. Further, as the increase was very small, it also may be due to a natural month-to-month variation as opposed to trending in the direction of increased sleep duration as observed in the other cities. Although Sweden did not initially institute the stringent COVID-19 containment measures other countries chose to enact (eg, shelter-in-place policies), the estimated sleep duration increased in Stockholm at a rate similar to that in other regions that did institute stringent containment policies at the time. This may reflect that people residing in Stockholm are following precautions to another

people worldwide, despite not being restricted to their homes. Finally, the observed increase in estimated sleep duration in New York City of >20 minutes could be explained in part by the high number of cases of COVID-19 in New York City and the associated strict stay-at-home policies.

The increase in the estimated sleep duration observed in this study in response to the COVID-19 pandemic of between 12 and 24 minutes is of similar magnitude to the approximately 25-minute increase in total sleep time induced by controlled-release zolpidem [22]. Thus, the COVID-19 mitigation policies are associated with an increase in estimated sleep duration comparable to the impact of a hypnotic medication. This may be clinically important, as sleep deficiency—which is highly prevalent in developed countries—increases susceptibility to viral infection and diminishes the immune response to vaccination [10,11].

Although we detected an increase in estimated sleep duration after the onset of COVID-19 in several different metropolitan areas, the amount of sleep disturbance may have changed after the onset of the pandemic. Previous research has shown that sleep quality suffers in the face of major crises [2]. Specifically, research has documented a significant increase in sleep difficulty and arousal from sleep amid the 2003 SARS outbreak [4], during the bombings in Israel during the Persian Gulf War [3], and in the aftermath of natural disasters [5,6], consistent with the reported increase in self-reported insomnia symptoms in Wuhan, China, at the start of the COVID-19 pandemic [23]. Our data indicating an increase in the estimated sleep duration do not preclude the possibility that more people were suffering from disturbed sleep. Additional studies administering validated questionnaires assessing insomnia symptoms as well as self-reported sleep quality would be required to address this question.

Limitations and Future Research

This study is subject to a number of limitations. (1) Although a strength of our study is the large number of respondents offering user-generated sleep data over time and across several geographic regions, a limitation is that we did not obtain data pertaining to sleep latency, awakenings from sleep, or time in bed after waking. (2) Although users do not commonly use the app for napping, we removed observations that were shorter than 60 minutes. Future research may examine naps and how napping behaviors changed due to the COVID-19 pandemic and its mitigation strategies. Further, we did not have access to sleep timing data, which would have been useful in understanding whether users were going to sleep earlier or sleeping later as a result of the COVID-19 pandemic. (3) Employment or income demographics, which are not collected by Sleep Cycle, may represent unmeasured, confounding variables in our study. For instance, individuals from higher socioeconomic groups were more likely to remain employed and to work from home during the COVID-19 pandemic [24]. Those who were furloughed or remained employed and were allowed to work from home may have had more flexible work hours and spent significantly less time commuting, allowing them to spend more time in bed [25]. Moreover, with the proliferation of low-cost delivery services (eg, food and other goods) in major urban centers, users who were able to work from home may have had increased available time for sleep [26]. These results may therefore not be generalizable to those working essential jobs, whose work hours and commuting time may have increased during the COVID-19 pandemic and who may be underrepresented in this sample. (4) The sleep tracker smartphone app is not validated, and because it does not account for waking after sleep onset, it is possible that the tracker overestimates sleep duration. (5) Those who download and use the sleep tracker smartphone app consistently may be more interested in sleep and health in general, which may represent selection bias, survivorship bias, and less generalizability to the general public.

Our study identifies several avenues for future research; some reflect the limitations of the app. First, it is important to explore whether the findings we observed are transient or will persist for an extended period and whether other aspects of sleep changed during the COVID-19 pandemic. Second, data from the Sleep Cycle app require validation using objective measures, such as polysomnography or actigraphy. Third, it would be informative to determine the prevalence of sleep disturbance or insomnia symptoms, such as difficulty falling asleep or difficulty maintaining sleep, among users of this app. Finally, an exploratory qualitative investigation with users of Sleep Cycle may help to determine if actual sleep duration in fact increased during the COVID-19 pandemic or if users experienced more sleep difficulties despite having increased time in bed.

Conclusion

Public health officials' policies to mitigate the COVID-19 pandemic have required profound changes in peoples' daily routines worldwide. Using approximately 2.9 million nights of objectively recorded sleep on 3 continents via a sleep tracking smartphone app, we observed an abrupt, significant increase in average estimated sleep duration among 8218 unique users in 5 cities in the months after the onset of the COVID-19 pandemic. Thus, not only have COVID-19 mitigation strategies reduced the potential of infection with SARS-CoV-2, but the resulting increase in sleep episode duration may also have improved the ability of the immune system to resist infection. Future research should explore whether this observation is transient or persists over the long term and whether other aspects of sleep have changed amid the COVID-19 pandemic.

Acknowledgments

All authors contributed to the conceptualization of this study. RR, MDW, and MA conducted the statistical analysis. This work was supported in part by gifts from Sleep Cycle and Alexandra Drane. RR and MA created the visuals. RR drafted the manuscript. All authors reviewed and approved the final manuscript. The authors thank Dr. Daniel Sääf from Sleep Cycle for aiding in the data acquisition.

Conflicts of Interest

The authors disclose all industry relationships in accordance with International Committee of Medical Journal Editors recommendations. RR has received fees from Denihan Hospitality, Rituals Cosmetics, Dagmejan, Asystem, and Sleep Cycle. SFQ reports personal fees from Jazz Pharmaceuticals and Best Doctors and is a consultant to Whispersom. MDW reports support from grants from the National Institute for Occupational Safety and Health (NIOSH) and the Brigham Research Institute and fees from the National Sleep Foundation and the University of Pittsburgh. MÉC reports personal fees from Vanda Pharmaceuticals and grants from the Australian-American Fulbright Commission. LKB reports consulting fees from University of Pittsburgh, Sygma, Insight, and Puget Sound Pilots. CAC reports grants to Brigham and Women's Hospital (BWH) from the National Heart, Lung, and Blood Institute, National Institute on Aging, NIOSH, and the Department of Defense; is or was a paid consultant to Bose, the Boston Celtics, the Boston Red Sox, Cephalon, the Institute of Digital Media and Child Development, the Klarman Family Foundation, Jazz Pharma, Merck, Purdue Pharma, Samsung, Teva Pharma Australia, the American Association of Retired Persons, the American Academy of Dental Sleep Medicine, the Eisenhower Medical Center, M. Davis and Company, Physician's Seal, the University of California San Diego, the University of Washington, the University of Michigan, the Maryland Sleep Society, the National Sleep Foundation, the Sleep Research Society, Tencent, and Vanda Pharmaceuticals, in which he also holds an equity interest. He also receives research and education support through BWH from Cephalon, Mary Ann & Stanley Snider via Combined Jewish Philanthropies, National Football League Charities, Jazz Pharmaceuticals Plc Inc, Philips Respironics Inc, Regeneron Pharmaceuticals, Teva Pharmaceuticals Industries Ltd, Sanofi SA, Optum, ResMed, San Francisco Bar Pilots, Sanofi, Schneider, Simmons, Sysco, Philips, and Vanda Pharmaceuticals; is or was an expert witness in legal cases, including those involving Advanced Power Technologies, Alvarado Hospital LLC, Amtrak, Bombardier Inc, C&J Energy Services, Casper Sleep Inc, Columbia River Bar Pilots, Complete General Construction Company, Dallas Police Department, Delta Airlines/Comair, Enterprise Rent-A-Car, Fédération des Médecins Résidents du Québec, FedEx, Greyhound Lines Inc/Motor Coach Industries/FirstGroup America, H.G. Energy LLC, Maricopa County, Arizona, Sheriff's Office, Murrieta Valley Unified School District, Pomerado Hospital, Palomar Health District, Puckett EMS, Purdue Pharma, South Carolina Central Railroad Company LLC, Steel Warehouse Inc, Union Pacific Railroad, United Parcel Service, and Vanda Pharmaceuticals. CAC serves as the incumbent of an endowed professorship provided to Harvard University by Cephalon, Inc, and receives royalties from McGraw Hill and Philips Respironics for the Actiwatch-2 and Actiwatch Spectrum devices. CAC's interests were reviewed and are managed by BWH and Partners HealthCare in accordance with their conflict of interest policies. All other authors have no conflicts to declare.

References

1. Solano JJ, Maki DG, Adirim TA, Shih RD, Hennekens CH. Public health strategies contain and mitigate COVID-19: a tale of two democracies. Am J Med 2020 Dec;133(12):1365-1366 [FREE Full text] [doi: 10.1016/j.amjmed.2020.08.001] [Medline: 32810463]
2. Lavie P. Sleep disturbances in the wake of traumatic events. N Engl J Med 2001 Dec 20;345(25):1825-1832. [doi: 10.1056/NEJMra012893] [Medline: 11752360]
3. Lavie P, Carmeli A, Mevorach L, Liberman N. Sleeping under the threat of the Scud: war-related environmental insomnia. Isr J Med Sci 1991;27(11-12):681-686. [Medline: 1757247]
4. Tsang HWH, Scudds RJ, Chan EYL. Psychosocial impact of SARS. Emerg Infect Dis 2004 Jul;10(7):1326-1327 [FREE Full text] [doi: 10.3201/eid1007.040090] [Medline: 15338536]
5. Mellman TA, David D, Kulick-Bell R, Hebding J, Nolan B. Sleep disturbance and its relationship to psychiatric morbidity after Hurricane Andrew. Am J Psychiatry 1995 Nov;152(11):1659-1663. [doi: 10.1176/ajp.152.11.1659] [Medline: 7485631]
6. Li X, Buxton O, Hikichi H, Haneuse S, Aida J, Kondo K, et al. Predictors of persistent sleep problems among older disaster survivors: a natural experiment from the 2011 Great East Japan earthquake and tsunami. Sleep 2018 Jul 01;41(7):3 [FREE Full text] [doi: 10.1093/sleep/zsy084] [Medline: 29726979]
7. Lewnard JA, Lo NC. Scientific and ethical basis for social-distancing interventions against COVID-19. Lancet Infect Dis 2020 Jun;20(6):631-633 [FREE Full text] [doi: 10.1016/S1473-3099(20)30190-0] [Medline: 32213329]
8. Krueger J, Majde J. Humoral links between sleep and the immune system: research issues. Ann N Y Acad Sci 2003 May;992(1):9-20. [doi: 10.1111/j.1749-6632.2003.tb03133.x] [Medline: 12794042]
9. Toda H, Williams JA, Gulledge M, Sehgal A. A sleep-inducing gene, nemuri, links sleep and immune function in Drosophila. Science 2019 Feb 01;363(6426):509-515 [FREE Full text] [doi: 10.1126/science.aat1650] [Medline: 30705188]
10. Cohen S, Doyle WJ, Alper CM, Janicki-Deverts D, Turner RB. Sleep habits and susceptibility to the common cold. Arch Intern Med 2009 Jan 12;169(1):62-67 [FREE Full text] [doi: 10.1001/archinternmed.2008.505] [Medline: 19139325]
11. Prather AA, Janicki-Deverts D, Hall MH, Cohen S. Behaviorally assessed sleep and susceptibility to the common cold. Sleep 2015 Sep 01;38(9):1353-1359 [FREE Full text] [doi: 10.5665/sleep.4968] [Medline: 26118561]
12. Lange T, Perras B, Fehm HL, Born J. Sleep enhances the human antibody response to hepatitis A vaccination. Psychosom Med 2003;65(5):831-835. [doi: 10.1097/01.psy.0000091382.61178.f1] [Medline: 14508028]
13. Spiegel K, Sheridan JF, Van Cauter E. Effect of sleep deprivation on response to immunization. JAMA 2002 Sep 25;288(12):1471-1472. [doi: 10.1001/jama.288.12.1471-a] [Medline: 12243633]

14. Hirshkowitz M, Whiton K, Albert SM, Alessi C, Bruni O, DonCarlos L, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. Sleep Health 2015 Mar;1(1):40-43. [doi: 10.1016/j.sleh.2014.12.010] [Medline: 29073412]
15. Czeisler Charles A. Glossary of standardized terminology for sleep–biological rhythm research. Sleep 1980 Sep;2(3):287-288. [doi: 10.1093/sleep/2.3.287]
16. CTA/NSF 2052.1-2016 (ANSI): Definitions and characteristics for wearable sleep monitors. American National Standards Institute. 2016. URL: https://webstore.ansi.org/standards/ansi/ctansf20522016ansi [accessed 2021-01-27]
17. Timeline: WHO's COVID-19 response. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline [accessed 2021-01-27]
18. Blume C, Schmidt MH, Cajochen C. Effects of the COVID-19 lockdown on human sleep and rest-activity rhythms. Curr Biol 2020 Jul 20;30(14):R795-R797 [FREE Full text] [doi: 10.1016/j.cub.2020.06.021] [Medline: 32693067]
19. Wright KP, Linton SK, Withrow D, Casiraghi L, Lanza SM, Iglesia HDL, et al. Sleep in university students prior to and during COVID-19 stay-at-home orders. Curr Biol 2020 Jul 20;30(14):R797-R798 [FREE Full text] [doi: 10.1016/j.cub.2020.06.022] [Medline: 32693068]
20. Woo J, Hyun S, Lee S, Kang S, Lee J, Kim L, et al. Productivity time lost by sleep disturbance among workers in Korea. J Korean Neuropsychiatr Assoc 2011 Jan;50(1):62-69.
21. Kim SJ, Lee YJ, Cho SJ, Cho IH, Lim W, Lim W. Relationship between weekend catch-up sleep and poor performance on attention tasks in Korean adolescents. Arch Pediatr Adolesc Med 2011 Sep 05;165(9):806-812. [doi: 10.1001/archpediatrics.2011.128] [Medline: 21893646]
22. Nowell P, Mazumdar S, Buysse D, Dew M, Reynolds C, Kupfer D. Benzodiazepines and zolpidem for chronic insomnia: a meta-analysis of treatment efficacy. JAMA 1997 Dec 24;278(24):2170-2177. [doi: 10.1001/jama.1997.03550240060035]
23. Li Y, Qin Q, Sun Q, Sanford LD, Vgontzas AN, Tang X. Insomnia and psychological reactions during the COVID-19 outbreak in China. J Clin Sleep Med 2020 Aug 15;16(8):1417-1418. [doi: 10.5664/jcsm.8524] [Medline: 32351206]
24. Pareek M, Bangash MN, Pareek N, Pan D, Sze S, Minhas JS, et al. Ethnicity and COVID-19: an urgent public health research priority. Lancet 2020 May 02;395(10234):1421-1422 [FREE Full text] [doi: 10.1016/S0140-6736(20)30922-3] [Medline: 32330427]
25. Czeisler M, Howard ME, Robbins R, Barger LK, Facer-Childs ER, Rajaratnam SM, et al. COVID-19: Public compliance with and public support for stay-at-home mitigation strategies. medRxiv Preprint posted online on April 24, 2020. [doi: 10.1101/2020.04.22.20076141]
26. De' R, Pandey N, Pal A. Impact of digital surge during Covid-19 pandemic: a viewpoint on research and practice. Int J Inf Manage 2020 Dec;55:102171 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102171] [Medline: 32836633]

Abbreviations

BWH: Brigham and Women's Hospital
NIOSH: National Institute for Occupational Safety and Health
SARS: severe acute respiratory syndrome

Edited by G Eysenbach; submitted 21.05.20; peer-reviewed by Y Senathirajah, A Prather, M Schmidt-Kraepelin; comments to author 13.07.20; revised version received 04.08.20; accepted 19.01.21; published 02.02.21.

Please cite as:
Robbins R, Affouf M, Weaver MD, Czeisler MÉ, Barger LK, Quan SF, Czeisler CA
Estimated Sleep Duration Before and During the COVID-19 Pandemic in Major Metropolitan Areas on Different Continents: Observational Study of Smartphone App Data
J Med Internet Res 2021;23(2):e20546
URL: https://www.jmir.org/2021/2/e20546
doi:10.2196/20546
PMID:33493126

Original Paper

Predicting Public Uptake of Digital Contact Tracing During the COVID-19 Pandemic: Results From a Nationwide Survey in Singapore

Young Ern Saw[1], BA; Edina Yi-Qin Tan[1], BA; Jessica Shijia Liu[1]; Jean CJ Liu[1,2], PhD

[1]Division of Social Sciences, Yale-NUS College, Singapore, Singapore

[2]Neuroscience and Behavioral Disorders Programme, Duke-NUS Medical School, Singapore, Singapore

Corresponding Author:
Jean CJ Liu, PhD
Division of Social Sciences
Yale-NUS College
28 College Avenue West #01-501
Singapore, 138533
Singapore
Phone: 65 6601 3694
Email: jeanliu@yale-nus.edu.sg

Abstract

Background: During the COVID-19 pandemic, new digital solutions have been developed for infection control. In particular, contact tracing mobile apps provide a means for governments to manage both health and economic concerns. However, public reception of these apps is paramount to their success, and global uptake rates have been low.

Objective: In this study, we sought to identify the characteristics of individuals or factors potentially associated with voluntary downloads of a contact tracing mobile app in Singapore.

Methods: A cohort of 505 adults from the general community completed an online survey. As the primary outcome measure, participants were asked to indicate whether they had downloaded the contact tracing app TraceTogether introduced at the national level. The following were assessed as predictor variables: (1) participant demographics, (2) behavioral modifications on account of the pandemic, and (3) pandemic severity (the number of cases and lockdown status).

Results: Within our data set, the strongest predictor of the uptake of TraceTogether was the extent to which individuals had already adjusted their lifestyles because of the pandemic (z=13.56; *P*<.001). Network analyses revealed that uptake was most related to the following: using hand sanitizers, avoiding public transport, and preferring outdoor over indoor venues during the pandemic. However, demographic and situational characteristics were not significantly associated with app downloads.

Conclusions: Efforts to introduce contact tracing apps could capitalize on pandemic-related behavioral adjustments among individuals. Given that a large number of individuals is required to download contact tracing apps for contact tracing to be effective, further studies are required to understand how citizens respond to contact tracing apps.

Trial Registration: ClinicalTrials.gov NCT04468581, https://clinicaltrials.gov/ct2/show/NCT04468581

(J Med Internet Res 2021;23(2):e24730) doi:10.2196/24730

KEYWORDS

contact tracing; COVID-19; mobile app; digital health; epidemiology

Introduction

Background

In May 2020, Google and Apple released the Exposure Notification System, which is an application programming interface that logs the following: who a phone user has been in contact with, for how long, and at what distance [1]. This release came 2 months after COVID-19 was declared a pandemic [2], allowing governments to identify and isolate contacts of confirmed cases through a process known as “contact tracing” [3,4].

Less than a year after the first reported cases, over 33 million individuals have tested positive for COVID-19 worldwide and more than 1 million have died [5]. To limit disease spread, over

half of the global population has been subjected to lockdowns involving school closures, workplace shutdowns, and movement restrictions [6]. Although these lockdowns are effective in tapering the epidemic curve [7], they are costly to the global economy and are unsustainable [8]. However, allowing the virus to spread unhindered could overwhelm the health care system and result in large-scale mortality [9,10].

To address both infection control and economic concerns, several countries have turned to contact tracing to keep the economy running [11,12]. Epidemiological modeling suggests that if (1) cases are effectively identified (through rigorous testing protocols), (2) contact tracing is comprehensive (identifying all possible exposure), and (3) contacts are quarantined in a timely manner, this strategy can curb the spread of the virus [4,11]. In an optimal scenario, 80% of contacts should be traced on the same day an individual tests positive [3,11].

Conventional Versus Digital Contact Tracing

Early during the pandemic (and in previous infectious disease outbreaks), contact tracing was manually performed [13]. Using a range of interview and surveillance techniques, a human contact tracer would typically identify an average of 36 contacts for each positive case [14]. Although this strategy allows for high levels of case detection when there are few cases [15], its labor-intensive format—requiring ~12 h of tracing for each positive case [16]—is difficult to scale up. Additionally, individuals who test positive may forget whom they have been in contact with, thus undermining the effectiveness of the process [11].

Considering these limitations of manual contact tracing, several mobile apps have been developed to facilitate automated contact tracing [17], for example, COVID Watch in the United States [18], COVIDSafe in Australia [19], and Corona-Warn-App in Germany [20]. These apps primarily track Bluetooth signals from phones in the vicinity [3], capturing contacts without the restraints of staffing or recall biases [4,11]. Further, phone apps can notify individuals swiftly after a contact tests positive, allowing them to be quickly isolated [3].

Understanding the Predictors of Uptake

Despite the potential of digital contact tracing, a recent meta-analysis concluded that owing to implementation barriers, manual contact tracing should remain the order of the day [12]. One major barrier pertains to the uptake of mobile apps. Several modelling studies have assessed parameters needed for the COVID-19 reproduction number (R_0) to fall below 1 [3,11,21]. R_0 refers to the number of infections spread from 1 positive case, and a value less than 1 indicates that the virus has been contained. For this to be achieved, contact tracing apps need to be downloaded by at least 56% of the population [21], which is much higher than the average rate of downloads globally (9%) [22].

To increase uptake, Qatar made it mandatory for residents to use the official contact tracing app [23]. Although this legislation led to high download rates (>90% [24]), the potential backlash from the public (eg, because of privacy concerns [25,26]) implies that few countries are likely to follow suit. Correspondingly, public health agencies would benefit from an understanding of the predictors of voluntary downloads [27], providing an empirical basis to nudge citizens and residents to voluntarily download contact tracing apps [28].

The Current Study

Given the urgent need to boost contact tracing apps, this study is the first to identify sociodemographic factors predicting voluntary uptake. Our study was conducted in Singapore, where the world's first nationwide contact tracing app TraceTogether was launched in March 2020 [29]. TraceTogether uses a centralized approach adopted by several governments [19]; namely, randomly generated user IDs are generated and shared via Bluetooth with phones in close proximity [30]. When individuals test positive for COVID-19, they consent to add both their own user IDs and those of their contacts to a centralized database. This is used to identify matches, and exposure notifications are then sent from the server to close contacts [31,32]. (As an alternative model, a decentralized approach could be used where both matches and notifications are made through the user's phone [33].)

As Singapore was the forerunner of this technology, the app has accrued 2.3 million users within 6 months, including approximately 40% of Singapore's resident population or 50% of all smartphone users (considering a smartphone penetration rate of 82%) [34,35]. Correspondingly, our study represents a "best case scenario" for app uptake after several months have elapsed. In terms of the epidemic curve, our study was conducted between April and July 2020, as the country was in a lockdown (April to May 2020). This period witnessed a peak in daily COVID-19 cases (April: >1000/day or 175 per million population), which gradually tapered over time (July: >100/day or 17.5 per million population).

Methods

Study Design and Population

Between April 3 and July 17, 2020, we recruited 505 adults who met the following eligibility criteria: (1) at least 21 years of age and (2) had lived in Singapore for a minimum of 2 years. All participants responded to online advertisements. Within the constraints of online sampling owing to the pandemic, we strove to obtain a representative sample by placing advertisements in a wide range of online community groups (eg, Facebook or WhatsApp groups among individuals in residential estates, universities, and workplaces) and by using paid online advertisements targeting the broad spectrum of Singapore residents.

Prior to study enrolment, participants provided informed consent in accordance with a protocol approved by the Yale-NUS College Ethics Review Committee (#2020-CERC-001; ClinicalTrials.gov ID NCT04468581). They then completed a 10-min online survey hosted on the platform Qualtrics [36]. Data were collected in accordance with the second phase of a larger study tracking COVID-19 responses, and findings from the first phase have been described previously [37,38].

Outcome Variable: Use of TraceTogether

As the primary outcome variable, participants were asked to indicate whether they had downloaded the government's contact tracing app TraceTogether (binary variables: 1=they had, 0=they had not).

Predictors

Demographics and Situational Variables

As predictors of TraceTogether usage, participants then reported the following demographic data: age, gender, citizenship, ethnicity, marital status, education level, house type (a proxy of socioeconomic status in Singapore), and household size. Based on the survey timestamp, we also included the following as predictors: (1) the total number of cases in Singapore to date, (2) whether the nation was in a lockdown at the time of participation (0=no, 1=yes), and (3) a self-reported measure of confidence the government could control COVID-19 spread (4-point scale: 1="not confident at all," 4="very confident").

Other Behavioral Modifications

As a basis of comparison, participants were also asked to identify which of 18 other behavioral modifications they had made as a result of the pandemic (apart from downloading TraceTogether). Specifically, participants were asked whether they had (1) washed their hands more frequently, (2) used hand sanitizers, (3) worn a mask in public voluntarily (before a law was passed), (4) avoided taking public transport, (5) stayed home more than usual, (6) avoided crowded places, (7) chosen outdoor over indoor venues, (8) missed or postponed social events, (9) changed their travel plans voluntarily, (10) reduced physical contact with others (eg, by not shaking hands), (11) avoided visiting hospitals or other health care settings, (12) avoided visiting places where COVID-19 cases had been reported, (13) maintained distance from people suspected of recent contact with a COVID-19–positive individual, (14) maintained distance from people who might have recently traveled to countries with an outbreak, (15) maintained distance from people with flu-like symptoms, (16) relied more on online shopping (eg, for groceries), (17) stocked up on more household supplies and groceries than usual, or (18) taken their children out of school (for each item, 0=the measure was not taken, 1=the measure was taken). These values were then summed to compute an aggregated measure of behavioral change (out of 18), and were included as a predictor to assess whether contact tracing usage was associated with conventional behavioral modifications one undertakes during an epidemic [39,40].

As part of the survey, participants were also asked to specify any other behavioral modifications (n=9, 1.8%) or no other behavioral modification (n=2, 0.4%). However, these data were excluded from the statistical analyses owing to the low base rate of affirmative responses.

Data Analysis Plan

For primary analysis, binary logistic regression was used to identify predictors TraceTogether uptake. In the first model (model 1), participants' demographics were included as predictors (age, citizenship, gender, marital status, education level, ethnicity, household type, and household size). Citizenship (base=others), gender (base=female), marital status (base=single), and ethnicity (base=Chinese) were coded as dummy variables. In the second model (model 2), we repeated the first model with the inclusion of situational variables (log-transformed total number of COVID-19 cases to date and lockdown status). Finally, in the third model (model 3), we repeated the second model with the inclusion of the total number of behavioral modifications as a predictor. All data were analyzed using SPSS (version 23, IBM Corp) and R (version 3.6.0, The R Foundation), with the type 1 familywise error rate controlled at α=.05 via Bonferroni correction (Bonferroni-adjusted α=.003 [.05/17 predictors]).

Results

Demographics of the Sample

Table 1 shows the wide range of demographic characteristics of our study cohort (N=505). Compared to the resident population, the sample was matched in the following characteristics: ethnic composition, household size, and housing type (a proxy of socioeconomic status in Singapore) (≤10% difference). However, compared to the resident population, the present participants were more likely to be female (n=313, 62.0% vs 51.1%), single or dating (n=234, 46.4% vs 31.6%), to have a higher level of education or no tertiary education (n=65, 12.9% vs 51.7%), and to be citizens of Singapore or of other countries (n=456, 90.3% vs 61.4%).

Table 1. Baseline characteristics of survey respondents (N=505).

Variable	Value
Age (years), mean (SD)	37.82 (11.31)
Number of behavioral modifications, mean (SD)	9.81 (3.82)
Gender, n (%)	
Female	313 (62.0)
Male	192 (38.0)
Citizenship, n (%)	
Singaporean	456 (90.3)
Others	49 (9.7)
Highest education, n (%)	
No formal education	2 (0.4)
Primary school	2 (0.4)
Secondary school	23 (4.6)
Junior college	26 (5.1)
Institution of Technical Education	12 (2.4)
Polytechnic (diploma)	88 (17.4)
University (degree)	265 (52.5)
Postgraduate (masters/PhD)	87 (17.2)
Ethnicity, n (%)	
Chinese	412 (81.6)
Malay	38 (7.5)
Indian	32 (6.3)
Eurasian	15 (3.0)
Others	8 (1.6)
Marital status, n (%)	
Single	170 (33.7)
Dating	64 (12.7)
Married	241 (47.7)
Widowed/separated/divorced	30 (5.9)
Household type, n (%)	
HDB[a] flat: 1-2 rooms	14 (2.8)
HDB flat: 3 rooms	50 (9.9)
HDB flat: 4 rooms	132 (26.1)
HDB flat: 5 rooms or executive flats	149 (29.5)
Condominium or private apartments	122 (24.2)
Landed property	38 (7.5)
Household size, n (%)	
1	26 (5.1)
2	88 (17.4)
3	119 (23.6)
4	133 (26.3)
5+	139 (27.5)

[a]HDB: Housing & Development Board.

Binary Logistic Regression

Of the 505 participants, 274 (54.3%; 95% CI 49.8%-58.7%) reported having downloaded TraceTogether. The download rate in this sample matches that of smartphone users in the resident population [34], and Table 2 describes the characteristics of users and nonusers.

Table 3 shows parameter estimates from logistic regression analyses of the predictors of TraceTogether uptake. No demographic or situational variable significantly predicted downloads (models 1 and 2). After controlling for these variables, the number of behavioral modifications emerged as a significant predictor (model 3); that is, with each unit increase in the number of behavioral modifications adopted, participants were 1.10 times more likely to download TraceTogether (z=13.56; P<.001).

Table 2. Characteristics of the users of TraceTogether (N=505).

Variable	TraceTogether usage	
	Users (n=274)	Nonusers (n=231)
Age (years), mean (SD)	38.57 (11.57)	36.95 (10.96)
Household type, mean (SD)	3.92 (1.18)	3.76 (1.21)
Household size, mean (SD)	3.54 (1.26)	3.53 (1.15)
Number of behavioral modifications, mean (SD)	10.33 (3.83)	8.96 (3.65)
Gender, n (%)		
Female	177 (64.6)	136 (58.9)
Male	97 (35.4)	95 (41.1)
Citizenship, n (%)		
Singaporean	240 (87.6)	216 (93.5)
Others	34 (12.4)	15 (6.5)
Highest education, n (%)		
No formal education	1 (0.4)	1 (0.4)
Primary school	1 (0.4)	1 (0.4)
Secondary school	15 (5.5)	8 (3.5)
Junior college	14 (5.1)	12 (5.2)
Institution of Technical Education	8 (2.9)	4 (1.7)
Polytechnic (diploma)	41 (15.0)	47 (20.3)
University (degree)	136 (49.6)	129 (55.8)
Postgraduate (masters/PhD)	58 (21.2)	29 (12.6)
Ethnicity, n (%)		
Chinese	218 (79.6)	194 (84.0)
Malay	20 (7.3)	18 (7.8)
Indian	19 (6.9)	13 (5.6)
Eurasian	13 (4.7)	2 (0.9)
Others	4 (1.5)	4 (1.7)
Marital status, n (%)		
Single	85 (31.0)	85 (36.8)
Dating	38 (13.9)	26 (11.3)
Married	133 (48.5)	108 (46.8)
Widowed/separated/divorced	18 (6.6)	12 (5.2)

Table 3. Logistic regression models of predictors of the uptake of TraceTogether (dependent variable=downloaded TraceTogether).

Variable	Model 1: demographics[a]		Model 2: demographics and situational variables		Model 3: demographics, situational variables, and behavioral modifications	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Age (years)	1.018 (1.00-1.04)	.09	1.020 (1.00-1.04)	.06	1.021 (1.00-1.04)	.05
Gender (base=female)	0.771 (0.53-1.12)	.17	0.799 (0.55-1.17)	.25	0.904 (0.61-1.34)	.61
Citizenship (base=others)	0.546 (0.26-1.14)	.11	0.597 (0.28-1.13)	.17	0.651 (0.30-1.39)	.27
Household type	1.082 (0.92-1.27)	.76	1.056 (0.90-1.25)	.52	1.011 (0.85-1.20)	.90
Household size	1.042 (0.89-1.23)	.62	1.028 (0.87-1.21)	.74	1.036 (0.88-1.22)	.67
Highest education	1.032 (0.89-1.19)	.67	1.029 (0.89-1.25)	.70	0.993 (0.85-1.16)	.92
Ethnicity (base=Chinese)						
Malay	1.057 (0.53-2.11)	.88	1.050 (0.52-2.14)	.89	0.980 (0.48-2.02)	.96
Indian	1.112 (0.52-2.37)	.78	0.984 (0.45-2.13)	.97	0.928 (0.43-2.03)	.85
Eurasian	3.454 (0.70-17.02)	.13	3.475 (0.70-17.37)	.13	3.402 (0.66-17.42)	.14
Others	0.720 (0.16-3.17)	.66	0.724 (0.16-3.29)	.68	0.851 (0.18-4.037)	.84
Marital status (base=single)						
Dating	1.505 (0.82-2.76)	.19	1.555 (0.84-2.90)	.16	1.392 (0.74-2.63)	.31
Married	0.968 (0.62-1.52)	.89	0.974 (0.61-1.55)	.91	0.900 (0.56-1.44)	.66
Widowed/separated/divorced	1.146 (0.47-2.79)	.76	1.155 (0.47-2.84)	.75	1.034 (0.41-2.59)	.94
Local COVID-19 cases to date (log)	N/A[b]	N/A	0.774 (0.50-1.21)	.26	0.752 (0.48-1.18)	.22
Lockdown (base=no lockdown)	N/A	N/A	0.561 (0.30-1.04)	.07	0.599 (0.32-1.12)	.11
Confidence in the government	N/A	N/A	1.372 (1.05-1.79)	.02	1.363 (1.04-1.79)	.03
Number of behavioral modifications	N/A	N/A	N/A	N/A	1.102 (1.05-1.16)	<.001[c]

[a]Model 1: Overall percentage of users correctly classified-56.6%, Nagelkerke R2-0.048; Model 2: Overall percentage of users correctly classified-58.2%, Nagelkerke R2-0.068; Model 3: Overall percentage of users correctly classified-60.2%, Nagelkerke R2-0.103.

[b]N/A: not applicable.

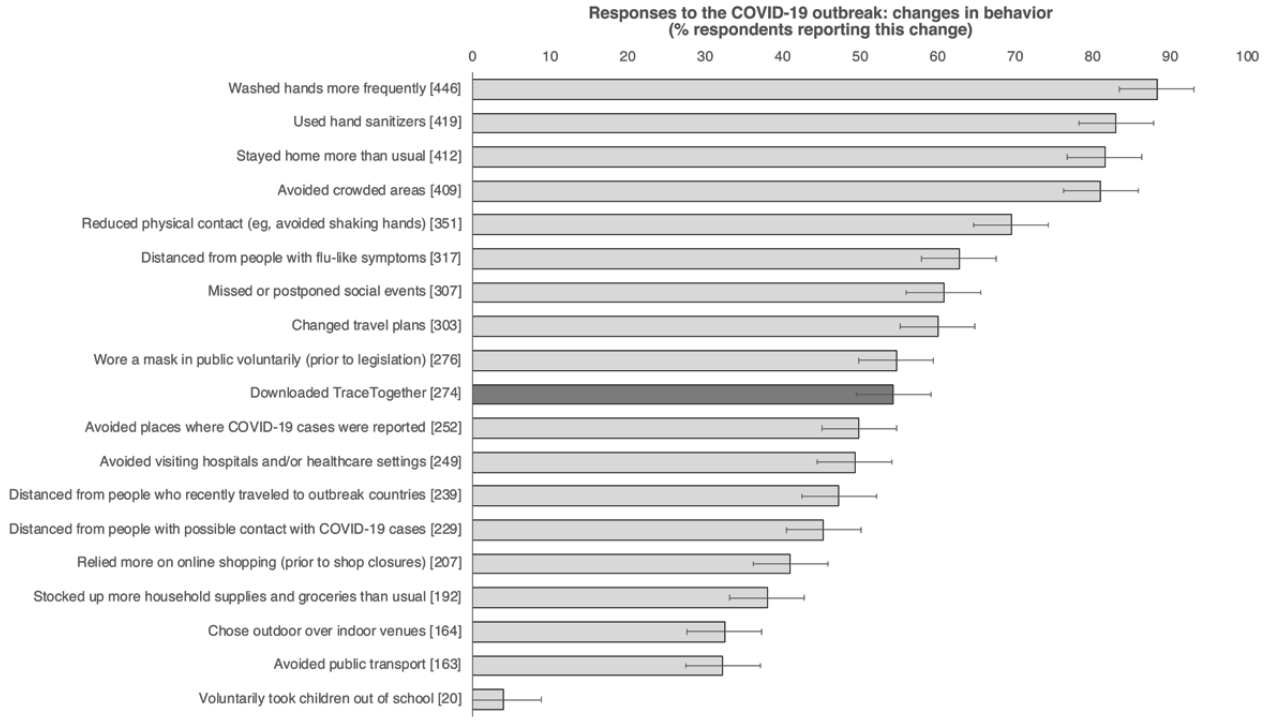
[c]*P*<.003 (following Bonferroni corrections).

Post Hoc Network Analysis

In the logistic regression analyses, TraceTogether downloads were predicted from the number of behavioral modifications because of the pandemic. To understand this association better, we conducted further exploratory analyses.

As shown in Figure 1, the majority of participants had modified their behaviors to curb the spread of COVID-19. The use of TraceTogether ranked 10th in the frequency of adoption (274/505, 54.3%), similar to the frequency of voluntary mask wearing (n=276, 54.7%).

Figure 1. Self-reported behavioral modifications , other than downloading TraceTogether, among the study participants undertaken in response to the COVID-19 outbreak in Singapore. Error bars=95% CI. Numbers in brackets represent the total number of respondents who reported the behavioral change.

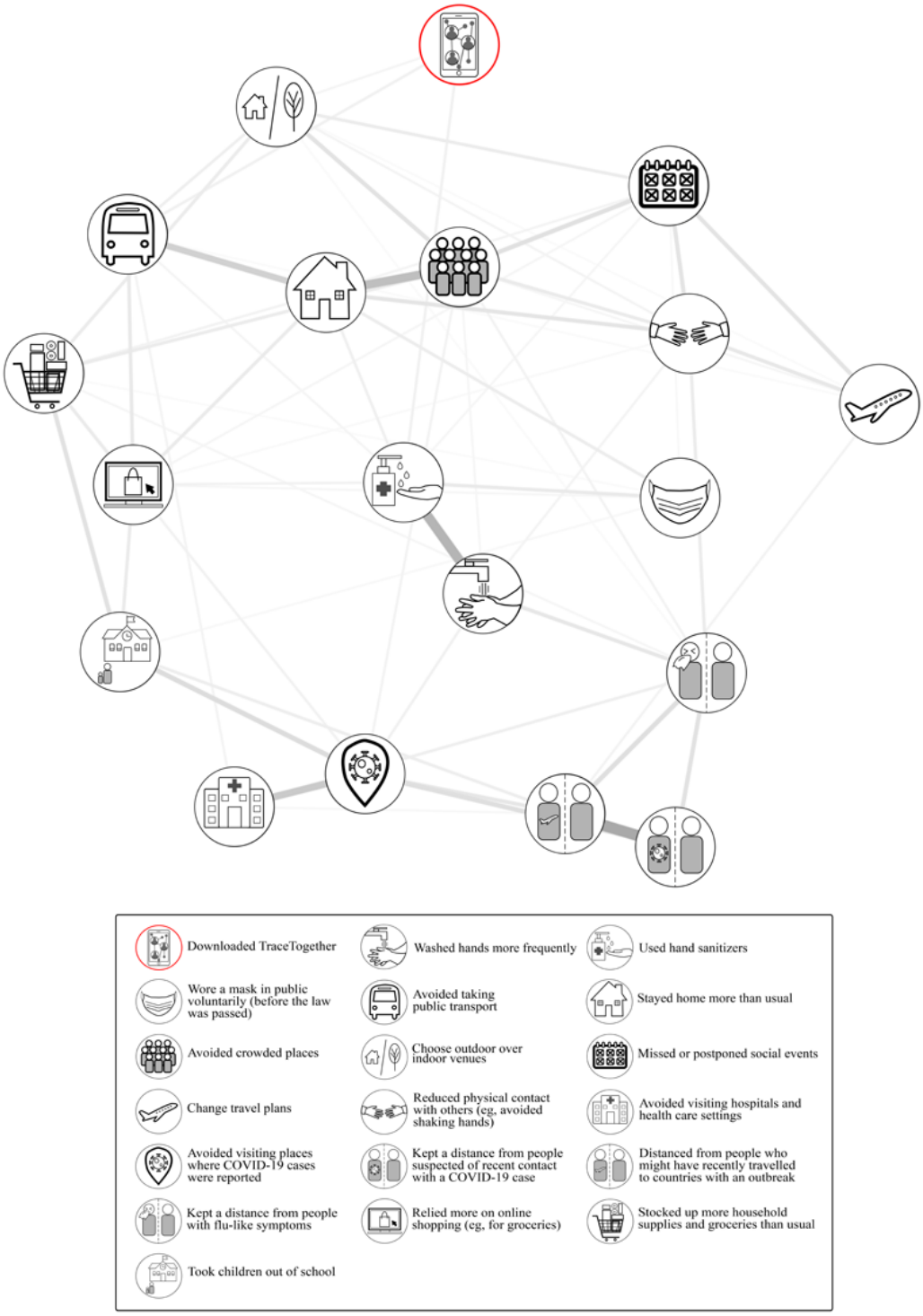


A corollary question is how TraceTogether usage is associated with other health protective behaviors; that is, how likely were people to download TraceTogether if they had modified their behavior in other ways? To address this question, we conducted network analyses by estimating a mixed graphical model (MGM) with the R package *mgm* [41]. MGM constructs weighted and undirected networks where the pathways among behaviors represent conditionally dependent associations, having controlled for the other associations in the network. Similar to partial correlations, each association (or "edge") is the average regression coefficient of two nodes. To avoid false-positive findings, we set small associations to 0 for the main models.

As shown in Figure 2, TraceTogether usage was associated with hand sanitizer use, avoidance of public transport, and a preference for outdoor vs indoor venues. The adjacency matrix (ie, numerical values for the average regression coefficient between two nodes) for Figure 2 is presented in Multimedia Appendix 1.

For sensitivity analysis, we performed logistic regression analysis using TraceTogether downloads as the dependent variable, and 18 other behavioral modifications (see Methods) as the predictors. Our conclusions did not change, as indicated in Multimedia Appendix 2.

Figure 2. A model depicting how TraceTogether usage relates to other pandemic-related behavioral changes. Line thickness represents the strength of an association.



Discussion

As lockdowns owing to COVID-19 ease globally, digital contact tracing will play an increasingly critical role in managing the epidemic curve. However, this requires the public to actively download a contact tracing app—a step that has proven elusive among public health agencies worldwide [12]. This study is the first to examine the demographic, behavioral, and situational factors potentially predicting the voluntary use of TraceTogether.

Behavioral Modifications

As our primary outcome, we observed that the number of behavioral modifications significantly predicted the use of TraceTogether. In other words, a person who had already changed his/her lifestyle on account of the pandemic was also

likely to download a contact tracing app. Network analyses revealed that downloads clustered with (1) using hand sanitizers, (2) avoiding public transport, and (3) preferring outdoor to indoor venues. This finding may suggest that public health campaigns could capitalize on other behavioral modifications when seeking to promote app downloads, for example, by printing information regarding a contact tracing app on the packaging of hand sanitizers or by framing the use of digital contact tracing as a preventive behavior. Policy makers might also expect app download rates to track behavioral modifications, anticipating, for example, higher download rates when the public fears an increase in COVID-19 cases (leading to more behavioral modifications) [42].

Theoretically, our findings further corroborate those of previous studies on how individuals change their behaviors during a pandemic. Based on prior outbreaks, a taxonomy of modifications had been identified whereby (1) "avoidant behaviors" are measures taken to avoid contact with potential carriers (eg, avoiding crowded places), while (2) "prevention behaviors" are those associated with maintaining hygiene (eg, regular hand washing) [42]. Extrapolating to the technological realm, our findings suggest that the use of a contact tracing app cuts across this taxonomy, since downloads were associated with *both* avoidant (avoiding public transport and preferring outdoor venues) and prevention behaviors (using hand sanitizers). Moving forward, we urge further studies to revise these classification systems in light of new technological developments.

Demographic and Situational Factors

Apart from behavioral modifications, it is notable that no demographic (eg, age, gender, etc) or situational variable (eg, number of COVID-19 cases and lockdown status) significantly predicted TraceTogether uptake. Prior to our study, it would have been conceivable that only a subset of the population would download a contact tracing app (eg, demographic groups based on gender, educational level, or age) [27,38,43]. By contrast, our findings highlight how uptake of digital contact tracing apps cuts across demographic groups.

While the lack of significant associations may be counterintuitive, a recent study reported similar results when predicting COVID-19–related behavioral modifications [42]. In a multinational survey, Harper et al [42] similarly observed that demographic and situational variables were unrelated to behavioral modifications owing to the pandemic. Since behavioral modifications predicted the use of a digital contact tracing app in our study, it seems reasonable to observe an analogous pattern here; that is, our results are unlikely to be based on false-negative outcomes.

As public health agencies develop strategies to promote downloads for contact tracing apps, the pattern of our findings may in turn suggest that demographic-specific messages are not needed. This is encouraging because the behavioral sciences offer widespread measures to "nudge" the general population [44]. In this case, the general public simply needs a one-off nudge to download the contact tracing app, after which the app functions independently in the background. Thus, if governments can nudge users in this first step (eg, by introducing incentives to download or by introducing contact tracing as an opt-out feature of existing government apps), it may be possible to attain the download rates necessary for contact tracing to be effective. Simultaneously, we urge further studies on the acceptance of such strategies; considering public concerns regarding privacy [25,26], any widespread intervention would need to be introduced cautiously.

Limitations

Our study has several limitations of note. As the first study of its kind, we made several choices at the exclusion of others. First, we opted for a cross-sectional design that precludes strong conclusions regarding causality. Second, we included an online sample to minimize person-to-person contact during the pandemic. Although we sampled individuals from a wide array of demographic groups, respondents were not representative of the general nationwide population; this may have deterred the establishment of potential associations among variables (eg, by including a high proportion of educated participants). Third, our survey relied on participants' self-reported use of a contact tracing app. Although our download rate is similar to that of the general population, further studies may seek to verify actual usage (eg, by incorporating survey questions in a contact tracing app). Fourth, our survey was not intended to measure every aspect of TraceTogether usage, and there were several notable omissions (eg, reasons why individuals chose to use or not use the app, phone ownership, and usage-related questions). Indeed, our model metrics (eg, small Nagelkerke R^2) indicate small effect sizes, highlighting the need for further studies to include a more comprehensive set of variables that may account for app downloads. Finally, we examined TraceTogether—an app with a centralized contact tracing protocol. Future studies are required to assess whether our findings extend to apps with decentralized protocols or to other forms of digital contact tracing that do not rely on mobile apps (eg, public acceptance of cloud-based contact in South Korea).

Conclusion

In conclusion, the potential contribution of digital technology to pandemic management is receiving increasing attention. What remains unclear, however, is how this technology is received and how best to promote its uptake. Focusing on contact tracing, this study shows that downloads of a mobile app was best predicted from the adoption of other infection control measures such as increased hand hygiene. In other words, the introduction of digital contact tracing is not merely a call to "trace together" but rather to "modify together," to use contact tracing apps as part of the broader spectrum of behavioral modifications during a pandemic.

Acknowledgments

This research was funded by a grant awarded to JCJL from the JY Pillay Global Asia Programme (grant number: IG20-SG002).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Adjacency matrix.
[PDF File (Adobe PDF File), 35 KB - jmir_v23i2e24730_app1.pdf]

Multimedia Appendix 2
Sensitivity analysis.
[PDF File (Adobe PDF File), 32 KB - jmir_v23i2e24730_app2.pdf]

References

1. Etherington D. Apple and Google launch exposure notification API, enabling public health authorities to release apps. TechCrunch. 2020 May 20. URL: http://techcrunch.com/2020/05/20/apple-and-google-launch-exposure-notification-api-enabling-public-health-authorities-to-release-apps/ [accessed 2020-09-30]
2. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2020-09-30]
3. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 Mar 31:eabb6936. [doi: 10.1126/science.abb6936] [Medline: 32234805]
4. Hellewell J, Abbott S, Gimma A, Bosse NI, Jarvis CI, Russell TW, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Lancet Glob Health 2020 Apr;8(4):e488-e496 [FREE Full text] [doi: 10.1016/S2214-109X(20)30074-7] [Medline: 32119825]
5. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-08-06]
6. Blanchard O, Philippon T, Pisani-Ferry J. A new policy toolkit is needed as countries exit COVID-19 lockdowns. Bruegel. 2020 Jun. URL: https://www.bruegel.org/wp-content/uploads/2020/06/PC-12-2020-230620.pdf [accessed 2020-08-06]
7. Lau H, Khosrawipour V, Kocbach P, Mikolajczyk A, Schubert J, Bania J, et al. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. J Travel Med 2020 May 18;27(3):taaa037 [FREE Full text] [doi: 10.1093/jtm/taaa037] [Medline: 32181488]
8. Fernandes N. Economic Effects of Coronavirus Outbreak (COVID-19) on the World Economy. SSRN Journal 2020 Mar 23. [doi: 10.2139/ssrn.3557504]
9. Betsch C, Wieler LH, Habersaat K, COSMO group. Monitoring behavioural insights related to COVID-19. Lancet 2020 Apr 18;395(10232):1255-1256 [FREE Full text] [doi: 10.1016/S0140-6736(20)30729-7] [Medline: 32247323]
10. Legido-Quigley H, Asgari N, Teo YY, Leung GM, Oshitani H, Fukuda K, et al. Are high-performing health systems resilient against the COVID-19 epidemic? Lancet 2020 Mar 14;395(10227):848-850 [FREE Full text] [doi: 10.1016/S0140-6736(20)30551-1] [Medline: 32151326]
11. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. Lancet Public Health 2020 Aug;5(8):e452-e459 [FREE Full text] [doi: 10.1016/S2468-2667(20)30157-2] [Medline: 32682487]
12. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. Lancet Digit Health 2020 Nov;2(11):e607-e621 [FREE Full text] [doi: 10.1016/S2589-7500(20)30184-9] [Medline: 32839755]
13. Chen K, Li C, Lin J, Chen W, Lin H, Wu H. A feasible method to enhance and maintain the health of elderly living in long-term care facilities through long-term, simplified tai chi exercises. J Nurs Res 2007 Jun;15(2):156-164. [doi: 10.1097/01.jnr.0000387610.78273.db] [Medline: 17551897]
14. Keeling MJ, Hollingsworth TD, Read JM. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). J Epidemiol Community Health 2020 Oct;74(10):861-866 [FREE Full text] [doi: 10.1136/jech-2020-214051] [Medline: 32576605]
15. Niehus R, De Salazar PM, Taylor AR, Lipsitch M. Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers. medRxiv. Preprint posted online February 18, 2020 [FREE Full text] [doi: 10.1101/2020.02.13.20022707] [Medline: 32511442]
16. Rubin R. Building an “Army of Disease Detectives” to Trace COVID-19 Contacts. JAMA 2020 May 21:2357-2360. [doi: 10.1001/jama.2020.8880] [Medline: 32437502]
17. LI J, Guo X. Global Deployment Mappings and Challenges of Contact-tracing Apps for COVID-19. SSRN Journal 2020 May 26. [doi: 10.2139/ssrn.3609516]

18. Morgan AU, Balachandran M, Do D, Lam D, Parambath A, Chaiyachati KH, et al. Remote Monitoring of Patients with Covid-19: Design, implementation, and outcomes of the first 3,000 patients in COVID Watch. NEJM Catalyst 2020 Jul 21;1(4). [doi: 10.1056/CAT.20.0342]
19. COVIDSafe app. Australian Government: Department of Health. URL: https://www.health.gov.au/resources/apps-and-tools/covidsafe-app [accessed 2020-09-30]
20. Reelfs J, Hohlfeld O, Poese I. Corona-Warn-App: Tracing the start of the official COVID-19 exposure notification app for Germany. arXiv csCY Preprint posted online July 25, 2020 [FREE Full text]
21. Hinch R, Probert W, Nurtay A, Kendall M, Wymant C, Hall M, et al. Effective configurations of a digital contact tracing app: A report to NHSX. 2020 May. URL: https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf?1587531217 [accessed 2020-08-06]
22. Chan S. COVID-19 Contact tracing apps reach 9% adoption In most populous countries. SensorTower. 2020 Jul 14. URL: https://sensortower.com/blog/contact-tracing-app-adoption [accessed 2020-09-30]
23. Qatar makes COVID-19 app mandatory, experts question efficiency. Al Jazeera. 2020 May 26. URL: https://www.aljazeera.com/news/2020/05/qatar-covid-19-app-mandatory-experts-question-efficiency-200524201502130.html [accessed 2020-09-30]
24. Rivero N. Global contact tracing app downloads lag behind effective levels. Quartz. 2020 Jul 15. URL: https://qz.com/1880457/global-contact-tracing-app-downloads-lag-behind-effective-levels/ [accessed 2020-10-01]
25. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19 pandemic. Nat Med 2020 Apr;26(4):463-464 [FREE Full text] [doi: 10.1038/s41591-020-0832-5] [Medline: 32284619]
26. Anderson M, Auxier B. Most Americans don't think cellphone tracking will help limit COVID-19, are divided on whether it's acceptable. Pew Research Center. 2020 Apr 16. URL: https://www.pewresearch.org/fact-tank/2020/04/16/most-americans-dont-think-cellphone-tracking-will-help-limit-covid-19-are-divided-on-whether-its-acceptable/ [accessed 2020-09-30]
27. Thorneloe R, Epton T, Fynn W, Daly M, Stanulewicz N, Kassianos A, et al. Scoping review of mobile phone app uptake and engagement to inform digital contact tracing tools for COVID-19. PsyArXiv Preprint posted online May 07, 2020. [doi: 10.31234/osf.io/qe9b6]
28. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: Cross-country survey evidence. JMIR Mhealth Uhealth 2020 Jul 24:e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
29. How Singapore's Covid-19 contact tracing app drew inspiration from a US high school project. Reuters. 2020 Jun 10. URL: https://www.scmp.com/news/asia/southeast-asia/article/3088389/how-singapores-covid-19-contact-tracing-app-drew [accessed 2020-09-30]
30. Bluetooth tracking and COVID-19: A tech primer. Privacy International. 2020 Mar 31. URL: https://www.privacyinternational.org/explainer/3536/bluetooth-tracking-and-covid-19-tech-primer [accessed 2020-10-02]
31. Bay J, Kek J, Tan A, Hau C, Yongquan L, Tan J, et al. BlueTrace: A privacy-preserving protocol for community-driven contact tracing across borders. Government Technology Agency-Singapore Tech Rep. 2020. URL: https://pdfs.semanticscholar.org/b460/fbaf2041cec41ee1266fcf1b60d3683d137b.pdf [accessed 2020-08-06]
32. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Digital contact tracing: comparing the capabilities of centralised and decentralised data architectures to effectively suppress the COVID-19 epidemic whilst maximising freedom of movement and maintaining privacy. University of Oxford. 2020 May 07. URL: https://tinyurl.com/y4bjggke [accessed 2020-08-06]
33. Cho H, Ippolito D, Yu YW. Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. arXiv csCR. e-Print posted online on March 25, 2020 2020 Mar 25 [FREE Full text]
34. TraceTogether. Government of Singapore. 2020 Jul 20. URL: https://www.tracetogether.gov.sg [accessed 2020-08-06]
35. Smartphone penetration rate as share of the population in Singapore from 2015 to 2025. Statista. 2020 Jul 30. URL: https://www.statista.com/statistics/625441/smartphone-user-penetration-in-singapore/ [accessed 2020-12-01]
36. Liu J. COVID-19 in Singapore: A nation-wide survey. OFSH. 2020 Jun 24. URL: https://osf.io/pv3bj/ [accessed 2020-10-01]
37. Liu JCJ, Tong EMW. The Relation Between Official WhatsApp-Distributed COVID-19 News Exposure and Psychological Symptoms: Cross-Sectional Survey Study. J Med Internet Res 2020;22(9):e22142 [FREE Full text] [doi: 10.2196/22142] [Medline: 32877349]
38. Long V, Liu J. Behavioral changes during the COVID-19 pandemic: Results of a national survey in Singapore. medXriv Preprint posted online on August 07, 2020. [doi: 10.1101/2020.08.06.20169870]
39. Bali S, Stewart KA, Pate MA. Long shadow of fear in an epidemic: fearonomic effects of Ebola on the private sector in Nigeria. BMJ Glob Health 2016;1(3):e000111 [FREE Full text] [doi: 10.1136/bmjgh-2016-000111] [Medline: 28588965]
40. Vinck P, Pham PN, Bindu KK, Bedford J, Nilles EJ. Institutional trust and misinformation in the response to the 2018-19 Ebola outbreak in North Kivu, DR Congo: a population-based survey. Lancet Infect Dis 2019 May;19(5):529-536. [doi: 10.1016/S1473-3099(19)30063-5] [Medline: 30928435]
41. Haslbeck JMB, Waldorp LJ. mgm: Estimating Time-Varying Mixed Graphical Models in High-Dimensional Data. arXiv Preprint posted online October 23, 2015. [doi: 10.1111/j.1750-4910.2016.tb00231.x]

42. Harper CA, Satchell LP, Fido D, Latzman RD. Functional Fear Predicts Public Health Compliance in the COVID-19 Pandemic. Int J Ment Health Addict 2020 Apr 27:1-14 [FREE Full text] [doi: 10.1007/s11469-020-00281-5] [Medline: 32346359]
43. Karekla M, Kasinopoulos O, Neto DD, Ebert DD, Van Daele T, Nordgreen T, et al. Best Practices and Recommendations for Digital Interventions to Improve Engagement and Adherence in Chronic Illness Sufferers. European Psychologist 2019 Jan;24(1):49-67. [doi: 10.1027/1016-9040/a000349]
44. Reisch LA, Sunstein CR. Do Europeans Like Nudges? SSRN Journal 2020 Jan 26;11(4):310-325. [doi: 10.2139/ssrn.2739118]

Abbreviations

MGM: mixed graphical model
R_0: reproduction number

Edited by G Eysenbach; submitted 02.10.20; peer-reviewed by T McCall, J Li, N Guo; comments to author 03.11.20; revised version received 01.12.20; accepted 15.01.21; published 03.02.21.

Please cite as:
Saw YE, Tan EYQ, Liu JS, Liu JCJ
Predicting Public Uptake of Digital Contact Tracing During the COVID-19 Pandemic: Results From a Nationwide Survey in Singapore
J Med Internet Res 2021;23(2):e24730
URL: https://www.jmir.org/2021/2/e24730
doi:10.2196/24730
PMID:33465034

Original Paper

Surveillance Metrics of SARS-CoV-2 Transmission in Central Asia: Longitudinal Trend Analysis

Lori Ann Post[1], PhD; Elana T Benishay[2], BSc; Charles B Moss[3], PhD; Robert Leo Murphy[4], MD; Chad J Achenbach[5], MD, MPH; Michael G Ison[5], MD, MSc; Danielle Resnick[6], PhD; Lauren Nadya Singh[1], MPH; Janine White[1], MA; Azraa S Chaudhury[2], BA; Michael J Boctor[2], BSc; Sarah B Welch[1], MPH; James Francis Oehmke[1], PhD

[1]Buehler Center for Health Policy and Economics, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Institute of Food and Agricultural Sciences, University of Florida, Gainsville, FL, United States

[4]Institute for Global Health, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[5]Divison of Infectious Disease, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[6]International Food Policy Research Institute, Washington, DC, United States

Corresponding Author:
Lori Ann Post, PhD
Buehler Center for Health Policy and Economics
Feinberg School of Medicine
Northwestern University
420 E Superior
Chicago, IL, 60611
United States
Phone: 1 203 980 7107
Email: lori.post@northwestern.edu

Abstract

Background: SARS-CoV-2, the virus that caused the global COVID-19 pandemic, has severely impacted Central Asia; in spring 2020, high numbers of cases and deaths were reported in this region. The second wave of the COVID-19 pandemic is currently breaching the borders of Central Asia. Public health surveillance is necessary to inform policy and guide leaders; however, existing surveillance explains past transmissions while obscuring shifts in the pandemic, increases in infection rates, and the persistence of the transmission of COVID-19.

Objective: The goal of this study is to provide enhanced surveillance metrics for SARS-CoV-2 transmission that account for weekly shifts in the pandemic, including speed, acceleration, jerk, and persistence, to better understand the risk of explosive growth in each country and which countries are managing the pandemic successfully.

Methods: Using a longitudinal trend analysis study design, we extracted 60 days of COVID-19–related data from public health registries. We used an empirical difference equation to measure the daily number of cases in the Central Asia region as a function of the prior number of cases, level of testing, and weekly shift variables based on a dynamic panel model that was estimated using the generalized method of moments approach by implementing the Arellano-Bond estimator in R.

Results: COVID-19 transmission rates were tracked for the weeks of September 30 to October 6 and October 7-13, 2020, in Central Asia. The region averaged 11,730 new cases per day for the first week and 14,514 for the second week. Infection rates increased across the region from 4.74 per 100,000 persons to 5.66. Russia and Turkey had the highest 7-day moving averages in the region, with 9836 and 1469, respectively, for the week of October 6 and 12,501 and 1603, respectively, for the week of October 13. Russia has the fourth highest speed in the region and continues to have positive acceleration, driving the negative trend for the entire region as the largest country by population. Armenia is experiencing explosive growth of COVID-19; its infection rate of 13.73 for the week of October 6 quickly jumped to 25.19, the highest in the region, the following week. The region overall is experiencing increases in its 7-day moving average of new cases, infection, rate, and speed, with continued positive acceleration and no sign of a reversal in sight.

Conclusions: The rapidly evolving COVID-19 pandemic requires novel dynamic surveillance metrics in addition to static metrics to effectively analyze the pandemic trajectory and control spread. Policy makers need to know the magnitude of transmission

rates, how quickly they are accelerating, and how previous cases are impacting current caseload due to a lag effect. These metrics applied to Central Asia suggest that the region is trending negatively, primarily due to minimal restrictions in Russia.

(J Med Internet Res 2021;23(2):e25799) doi:10.2196/25799

KEYWORDS

SARS-CoV-2 surveillance; second wave; wave two; global COVID-19 surveillance; Central Asia public health surveillance; Central Asia COVID-19; Central Asia surveillance metrics; dynamic panel data; generalized method of moments; Central Asia econometrics; Central Asia SARS-CoV-2; Central Asia COVID-19 surveillance system; Central Asia COVID-19 transmission speed; Central Asia COVID transmission acceleration; COVID-19 transmission deceleration; COVID-19 transmission jerk; COVID-19 7-day lag; SARS-CoV-2; Arellano-Bond estimator, generalized method of moments; GMM; Armenia; Azerbaijan; Cyprus; Faeroe Islands; Georgia; Gibraltar; Kazakhstan; Kosovo; Kyrgyzstan; Macedonia; Russia; Tajikistan Turkey; Turkmenistan; Uzbekistan; COVID-19; surveillance; longitudinal; trend; trend analysis; monitoring; public health; infectious disease; transmission; risk; management; policy; prevention

Introduction

Background

On December 29, 2019, 4 cases of “pneumonia of unknown etiology” were reported in Wuhan, Hubei Province, China [1]. What began as 4 isolated cases escalated into a global pandemic of SARS-CoV-2, the virus that causes COVID-19. To date, the caseload has reached 68,645,081 confirmed cases, and 1,564,496 deaths have been confirmed globally [2]. Nations have been greatly impacted by the pandemic, which has resulted in significant morbidity and mortality, food insecurity, and an economic recession that has not bottomed out. Global leaders are struggling to balance disease control with salvaging their plummeting economies in the face of a global pandemic [3]. This study aims to examine where and when SARS-CoV-2 was transmitted in Central Asia within the context of a global pandemic by delving into the environmental, sociocultural, and public health characteristics of COVID-19. For the purposes of this study, we use the World Bank’s definition of Central Asia, which includes Armenia, Azerbaijan, Cyprus, Faeroe Islands, Georgia, Gibraltar, Kazakhstan, Kosovo, Kyrgyzstan, Macedonia, Russia, Tajikistan, Turkey, Turkmenistan, and Uzbekistan [4].

History of Central Asia

Central Asia is largely composed of nation states that are former Soviet Union member countries. The Union of Soviet Socialist Republics (USSR) was dissolved in 1991 after controlling the region for 68 years, following a coup d’état during President Gorbachev’s administration. The former Soviet Union left a lasting legacy in the former national republics [5]. The USSR shifted from a centralized government and systems to independent states. Many essential institutions, such as national currency systems and military forces, were developed from the ground up. Many of these fledgling nations failed to achieve democracies, which led to more chaos and conflict. The “colored revolutions” from 1999-2005, a series of mass protests and riots, lead to the overthrow of the semiauthoritarian regimes in Serbia, Georgia, Ukraine, and Kyrgyzstan [6].

Public health systems in these new nations in Central Asia faced many challenges, including endemic infectious diseases [7] such as tuberculosis, HIV, and substance use disorders [8,9]. Alcohol poisoning is relatively common in the former USSR [10], contributing to low life expectancy in Russia [11,12]. Central Asia has one of the highest prevalence rates of tobacco misuse [13]. Misuses of alcohol, tobacco, and other substances represent both direct and indirect COVID-19 risk factors. Drinking alcohol, smoking, and misusing substances increases a person’s risk of being infected with SARS-CoV-2 and having worse outcomes if infected [14-17]. Moreover, persons who are self-isolating are at higher risk of substance use disorders [18-20]. Substance use disorders also contribute to increases chronic disease [21-23]. During past epidemics, reduced access to medical care for individuals with serious illness, such as HIV or tuberculosis, resulted in more deaths from complications due to a lack of health care resources [24,25].

Food and Water Security

Progress in food security [26] has stalled in Central Asia in recent years, with growth in malnourished and undernourished populations [27]. Malnourishment is a function of poverty [28,29].

Food insecurity is linked to environmental conditions caused by overuse of the Aral Sea, which has been depleted by 90% since 1960 to irrigate large areas of land [30,31]. Safe water is not available for 22 million people (31% of the population) throughout Kazakhstan, Kyrgyzstan, Uzbekistan, Tajikistan, and Turkmenistan. The majority of the affected individuals live in rural areas, where there are limited sewer connections and septic tanks [32]. People living near the highly polluted Aral Sea have higher levels of tuberculosis, anemia, and cancer, and they may be at higher risk of debilitating SARS-CoV-2 infection [33].

Beyond access to food, obesity is becoming more pervasive [34] as Central Asian foods mirror the “Western diet” of high fat and low grains [26]. Traditional dishes have high contents of sodium, likely due to the “Silk Road” pattern in which countries along the former Silk Road, a trade route through Central Asia established in the second century BC [35,36], use large quantities of salt for food preservation [37]. Obesity also contributes to chronic metabolic diseases and is associated with worse outcomes in those infected with SARS-CoV-2 [38].

Current Politics

Because Kazakhstan and China share a border, preventative measures in Kazakhstan were established as early as January

6, 2020, enforcing increased border sanitation and monitoring arrivals from China [39]. The first cases of COVID-19 in Kazakhstan, discovered in people arriving from Germany and Italy, were recorded on March 13. A state of emergency was announced on March 16; schools converted to remote learning, and quarantines were established in some areas as early as March 19 [39].

Many countries in the region have had disruptions in the labor market. The oil economies in Kazakhstan and Azerbaijan have been negatively impacted [40]. Tajikistan has experienced rising unemployment [29], and the poverty rate has risen across Central Asia [41-46].

State governments have attempted to minimize the impact of the virus; complaints about lack of adequate personal protective equipment in Russia were suppressed [47], medical authorities who gave public health advice in Turkey were criminally investigated [48], and the oppressive Turkmenistan government denied the existence of even a single case of SARS-CoV-2 [49].

In the electoral authoritarian regime of Azerbaijan, lockdown was established immediately following the first confirmed case of SARS-CoV-2 on February 28. The capital of Azerbaijan instated highly restrictive rules and shut down its border with the Islamic Republic of Iran, where cases were spreading quickly, the next day [50].

In Azerbaijan, violations of the quarantine mandate were punishable by fines, custodial restraint, and prison time [51]. Severe rules were briefly in place; people aged over 65 years were prohibited from leaving their homes [50], and citizens were required to send an SMS text message to a government telephone number to request permission to leave their home for up to three hours. The government inactivated the cellular service of political rivals so that they could not request permission to leave their residences and thus were confined to their homes [51]. The pandemic has been exploited as a means to restrict individual human rights in other countries as well; government access to private cell phone data was allowed in Armenia [52], and protests were restricted in Russia [53]. In both Uzbekistan and Tajikistan [54], penalties were introduced for the spread of false information about the virus through the media.

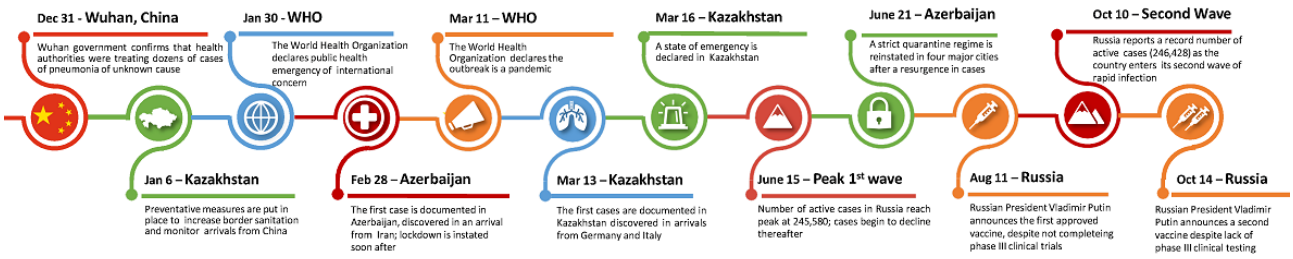
While Azerbaijan initially slowed the spread of the virus [55] and relaxed restrictions in May, the strict quarantine regime was reinstated in four major cities on June 21, 2020, after a surge in cases.

Shocking the international community, Russian President Vladimir Putin announced on August 11 that their country's health regulator was the first in the world to approve a SARS-CoV-2 vaccine for mass use [56]. This approval has been criticized as unsafe because the standard phase III clinical trials for new drugs were not completed at the time of his announcement [57]. While Russia's first approved vaccine, known as "Sputnik V," was still undergoing phase III trials, on October 14, President Putin announced the approval of a second SARS-CoV-2 vaccine [58]. Without adequate testing, we do not know the possible detrimental side effects of these vaccines, which may undermine attempts by the international community to provide safe immunization [56].

Other than vaccine efforts, Russia has not implemented significant public health measures. During the first wave of infections in March 2020, no financial support was given to small or medium-sized businesses despite instructions for employees to stay home on paid leave [59]. However, in mid-April, it was announced that over 100 billion rubles (US $1.325 billion) would be allocated to support small and medium-sized businesses [54]. The number of active cases in Russia reached a peak of 245,580 on June 15 and then began to decline [60]. With the second wave of infections in October after loosening restrictions, Russia saw a spike in cases and reported the fourth highest number of SARS-CoV-2 cases worldwide, behind the United States, India, and Brazil [61]. Russia reported a record number of 29,039 active cases on December 6 [60]. Despite record numbers of daily cases, with 26,097 new cases on December 8 for a total of 2,541,199 cases, protective measures continued to be rolled back. While gloves and masks are required in the Moscow metro, citizens must register their telephone number before entering a bar or nightclub; moreover, museums are closed, international flights are gradually being reinstated, and students in first through fifth grade are returning to school [62]. As early as July, the mayor of Moscow announced that wearing masks would no longer be required outdoors as new COVID-19 cases dwindled in the capital [54]. Figure 1 shows the timeline of the COVID-19 pandemic from December 2019 to October 2020 in Central Asia.

Without an effective vaccine to prevent COVID-19, Central Asian leaders require an effective SARS-CoV-2 surveillance system that enables their governments to make safe and informed decisions [63-70]. Public health departments [71-76] plus several universities [77] and media outlets [78,79] are tracking the novel coronavirus using raw data regarding the number of new infections, testing, positivity, basic reproduction number (R_0), and deaths, among other measures, such as local hospital capacity.

Figure 1. Timeline of the COVID-19 pandemic in Central Asia.



To that end, the objective of our research was to use a longitudinal trend analysis study design in concert with dynamic panel modeling and method of moments approaches to correct for existing surveillance data limitations [80,81]. Specifically, we measured significant weekly shifts in the increase, decrease, or plateaued transmission of SARS-CoV-2. Our study measured the underlying causal effect from the previous week that persisted through the current week, with a 7-day persistence rate to explain a clustering or declustering effect. The 7-day persistence represents an underlying disease transmission wave, where a large number of transmissions that resulted in a large number of infections on 1 day then "echoes" forward into a large number of new transmissions and hence a large number of new cases in the next 7 days. An example of the 7-day lag would be large sporting events in the United Kingdom that drew large crowds over an extended period of time even after cases were confirmed in the country. Other potential "super-spreader" events occurred in Turkmenistan, when a mass cycling rally was held on April 7 to celebrate World Health Day [82]. In summary, we measured negative and positive shifts in the transmission of SARS-CoV-2 or its acceleration and deceleration rates. We measured negative and positive shifts in the transmission of SARS-CoV-2 as well as the speed, acceleration or deceleration, and jerk rates along with the 7-day persistence, which do not suffer from sampling bias. For details, see Oehmke et al [80,81]. Our surveillance metric will provide public health surveillance data to inform governments that are making decisions regarding disease control, mitigation strategies, and reopening policies as they continue to manage this unprecedented situation.

Methods

This study relies on a longitudinal trend analysis of data collected from the Foundation for Innovative New Diagnostics (FIND) [83]. FIND compiles data from multiple sources across individual websites, statistical reports, and press releases. Data for the most recent 8 weeks were accessed from a GitHub repository that compiles data from multiple sources on the web; data for the most recent 4 weeks were accessed from the GitHub repository [84]. This resulted in a panel of 14 countries in Central Asia with 47-50 days in each panel (N=696). An empirical difference equation was specified in which the number of new positive cases in each country at each day is a function of the prior number of cases, the level of testing, and weekly shift variables that measure whether the contagion was growing faster than, the same as, or slower than in the previous weeks. This resulted in a dynamic panel model that was estimated using the generalized method of moments approach by implementing the Arellano-Bond estimator in R (R Project) [85].

Arellano-Bond estimation of difference equations has several statistical advantages: (1) it enables statistical examination of the predictive ability of a model and the validity of the model specification; (2) it corrects for autocorrelation and heteroscedasticity; (3) it has good properties for handling data with a small number of time periods and large number of states; and (4) it corrects for omitted variables and provides a statistical test of correction validity. With these advantages, the method is applicable to ascertaining and statistically validating changes in the evolution of the pandemic within a period of ≤1 week, such as changes in the reproduction rate. See Oehmke et al [80,81] for a detailed discussion of the methods. Finally, we calculated these indicators to inform public health leaders of where to take corrective action at a local level. China enjoyed great success at controlling the pandemic by closing down smaller geographical regions, preserving the larger economy and preventing other adverse outcomes from a national quarantine.

Results

Country Regression Results

We analyzed the 12 countries that are included in the Central Asia region as defined by the World Bank. The results of the associated regression supporting the weekly surveillance metrics are captured in Table 1. The Wald statistic for regression was significant (χ^2_7=14,217, P<.001). The Sargan statistic for validity was insignificant (χ^2_{550}=9, P>.99) and failed to reject the validity of overidentifying restrictions.

As shown in Table 1, the 1-day lag coefficient is positive and significant (1.075, P<.001), suggesting a clustering effect where the number of cases on a given day impact the number of cases on adjoining days. The 7-day lag coefficient and the impact of the limited testing over the weekend on case counts (the "weekend effect") are both insignificant. The shift parameters for the weeks of October 6 and 13 are also insignificant, suggesting that there were no major changes in the rate of disease transmission in the region between these 2 weeks in general. The coefficient for cumulative tests is insignificant.

Table 1. Arellano-Bond dynamic panel data modeling of the number of daily infections reported by country in Central Asia from September 30 to October 13, 2020.

Variable	Values	*P* value
1-day lag coefficient	1.075	<.001
7-day lag coefficient	–0.051	.89
Cumulative tests	0.000016	.55
Shift parameter, week of October 6	–0.157	.53
Shift parameter, week of October 13	–0.417	.12
Weekend effect[a]	–0.009	.99

[a]Weekend effect: impact of limited testing over the weekend on case counts.

Interpretation: Central Asian Regression Results

The 7-day lag and shift parameters suggest that there have been no recent changes in disease transmission rates. Additionally, there is no weekend effect or cumulative test effect.

Surveillance Results

Static and dynamic surveillance metrics for the weeks of October 6 and 13 are reflected in Tables 2-6, and Table 7 shows the most populous countries in Central Asia as of 2020. Tables 2 and 3 capture static metrics, including the number of new COVID-19 cases, number of cumulative COVID-19 cases, 7-day moving average of new cases, rate of infection, new deaths, cumulative deaths, 7-day moving average of number of deaths, and death rates. Novel dynamic metrics are reflected in Tables 4 and 5. These metrics include (1) speed, or the weekly average of new daily cases per 100,000 persons, (2) acceleration, or the day-to-day change in speed, (3) jerk, or the week-over-week change in acceleration, 4) 7-day persistence effect, or the number of new cases per 100,000 persons reported on a particular day that are associated with new cases reported 7 days previously. These novel metrics enable analysis of the impact of previous cases on current cases and identification of potential changes of the pandemic trajectory in the future.

Table 2 reflects static surveillance metrics for the week of September 30 to October 6, and Table 3 reflects those metrics for the week of October 7-13. The region averaged 11,730 new cases per day for the period ending on October 6 and 14,514 for the period ending on October 13. The infection rate increased across the region from 4.74 per 100,000 persons to 5.66. This increase in infection rate was accompanied by a slight increase in death rate from 0.09 per 100,000 persons to 0.11. Up to October 13, the region had reported 2,040,812 cumulative COVID-19 cases.

Table 2. Static surveillance metrics for the week of September 30 to October 6, 2020.

Country	New COVID-19 cases, n	Cumulative COVID-19 cases, n	7-day moving average of new cases	Infection rate (per 100,000 persons)	New deaths, n	Cumulative deaths, n	7-day moving average of the death rate	Death rate (per 100,000 persons)
Armenia	406	53,083	454.57	13.73	6	990	4.57	0.20
Azerbaijan	143	40,931	116.00	1.43	2	600	1.43	0.02
Cyprus	29	1876	19.00	2.42	1	23	0.14	0.08
Georgia	549	9245	482.71	14.76	4	58	3.14	0.11
Kazakhstan	66	108,362	64.86	0.36	21	1746	3.00	0.11
Kosovo	34	15,889	45.00	1.89	2	635	1.43	0.11
Kyrgyzstan	164	47,799	182.43	2.54	0	1066	0.29	0
North Macedonia	223	19,096	187.14	10.70	8	768	4.43	0.38
Russia	11481	1,231,277	9835.57	7.95	184	21,559	157.57	0.13
Tajikistan	40	10,014	41.14	0.43	0	78	0.43	0
Turkey	1511	327,557	1469.29	1.81	55	8553	60.43	0.07
Uzbekistan	397	59,343	427.00	1.18	4	489	3.29	0.01

Table 3. Static surveillance metrics for the week of October 7-13, 2020.

Country	New COVID-19 cases, n	Cumulative COVID-19 cases, n	7-day moving average of new cases	Infection rate (per 100,000 persons)	New deaths, n	Cumulative deaths, n	7-day moving average of the death rate	Death rate (per 100,000 persons)
Armenia	745	57,566	640.43	25.19	6	1032	6.00	0.20
Azerbaijan	277	42,381	207.14	2.76	3	612	1.71	0.03
Cyprus	83	2130	36.29	6.92	0	25	0.29	0.00
Georgia	569	12,841	513.71	15.29	9	102	6.29	0.24
Kazakhstan	83	108,984	88.86	0.45	22	1768	3.14	0.12
Kosovo	98	16,345	65.14	5.46	1	649	2.00	0.06
Kyrgyzstan	343	49,871	296.00	5.31	2	1092	3.71	0.03
North Macedonia	80	21,193	299.57	3.84	3	800	4.57	0.14
Russia	13,690	1,318,783	12,500.86	9.48	240	22,834	182.14	0.17
Tajikistan	37	10,297	40.43	0.40	0	79	0.14	0.00
Turkey	1632	338,779	1603.14	1.96	62	8957	57.71	0.07
Uzbekistan	323	61,642	328.43	0.96	2	511	3.14	0.01

Table 4. Novel surveillance metrics for the week of September 30 to October 6, 2020.

Country	Speed[a]	Acceleration[b]	Jerk[c]	7-day persistence effect on speed[d]
Armenia	15.4	0.4	0.3	–0.6
Azerbaijan	1.2	0.1	0	–0.1
Cyprus	1.6	0	–0.1	–0.1
Georgia	13.0	0.9	–0.2	–0.4
Kazakhstan	0.4	0	0	0
Kosovo	2.5	–0.2	–0.1	–0.1
Kyrgyzstan	2.8	0	–0.2	–0.1
North Macedonia	9.0	0.8	0.5	–0.3
Russia	6.8	0.3	0.1	–0.3
Tajikistan	0.4	0	0.0	0
Turkey	1.8	0	0.0	–0.1
Uzbekistan	1.3	–0.1	0.0	–0.1

[a]Speed: daily positives per 100,000 persons (weekly average of new daily cases per 100,000 persons).

[b]Acceleration: day-to-day change in the number of positives per day (weekly average per 100,000 persons).

[c]Jerk: week-over-week change in acceleration per 100,000 persons.

[d]7-day persistence effect on speed: number of new cases per day per 100,000 persons.

Table 5. Novel surveillance metrics for the week of October 7-13, 2020.

Country	Speed[a]	Acceleration[b]	Jerk[c]	7-day persistence effect on speed[d]
Armenia	21.7	1.6	0.7	–0.8
Azerbaijan	2.1	0.2	0.2	–0.1
Cyprus	3.0	0.6	0.4	–0.1
Georgia	13.8	0.1	0.5	–0.7
Kazakhstan	0.5	0	0	0
Kosovo	3.6	0.5	0.3	–0.1
Kyrgyzstan	4.6	0.4	0.2	–0.1
North Macedonia	14.4	–1.0	–1.6	–0.5
Russia	5.2	0.1	0.0	–0.2
Tajikistan	8.7	0.2	0.0	–0.4
Turkey	0.4	0	0.0	0
Uzbekistan	1.9	0	0.0	–0.1

[a]Speed: daily positives per 100,000 persons (weekly average of new daily cases per 100,000 persons).

[b]Acceleration: day-to-day change in the number of positives per day (weekly average per 100,000 persons).

[c]Jerk: week-over-week change in acceleration per 100,000 persons.

[d]7-day persistence effect on speed: number of new cases per day per 100,000 persons.

Table 6. Comparison of 1-day persistence in the four countries in Central Asia with positive significant positive accelerations for the week of October 6, 2020.

Country	1-day persistence	
	Week of September 30	Week of October 6
Armenia	16.1	21.5
Georgia	13.0	16.5
North Macedonia	8.8	14.8
Russia	7.0	9.1

Table 7. Most populous countries in Central Asia as of 2020.

Country	Population as of 2020, n
Russia	145,953,632
Turkey	84,621,255
Uzbekistan	33,469,203
Kazakhstan	18,776,707

Russia and Turkey had the highest 7-day moving averages in the region, at 9836 and 1469, respectively, for the week of October 6, and 12,501 and 1603, respectively, for the week of October 13 (Tables 2 and 3). In terms of infection rate, accounting for population, Russia was at 7.95 per 100,000 persons and Turkey was at 1.81 for the week of October 06. The countries that had the highest infection rates in the region included Georgia, at 14.76 per 100,000 persons, and Armenia, at 13.73. For the following week, the infection rate in Armenia jumped to the highest in the region, at 25.19 per 100,000 persons, while Georgia had a very slight increase to 15.29. Kazakhstan had the lowest infection rate in the region, at 0.36 per 100,000 persons for the week of October 6.

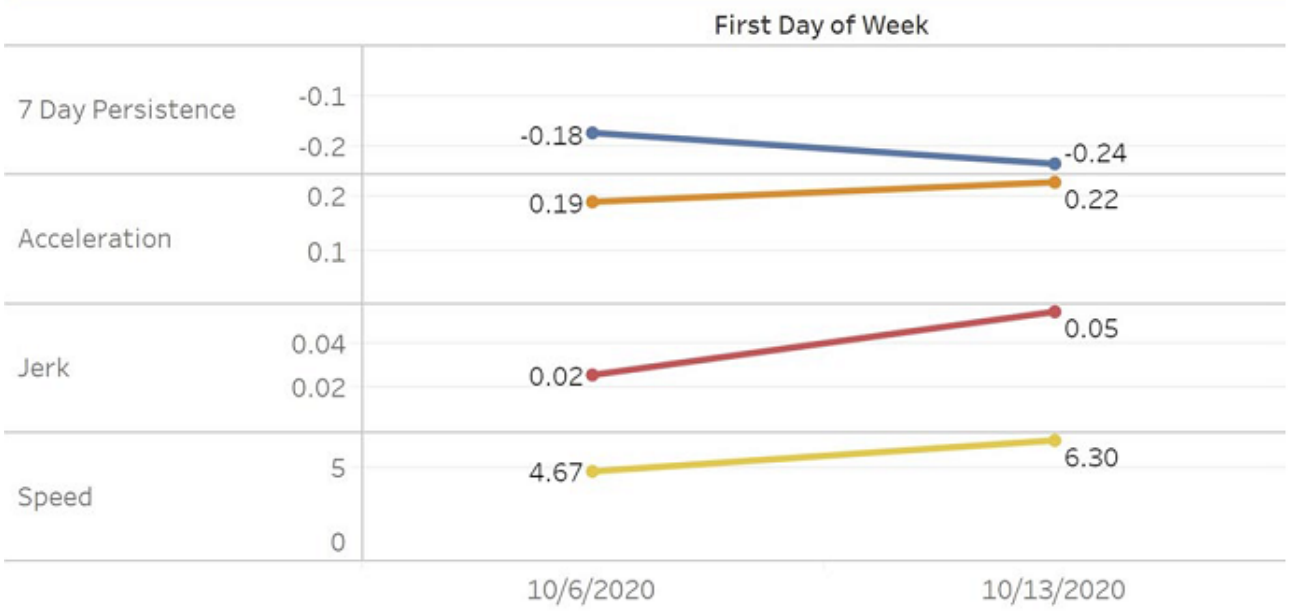
Russia and Turkey also had the highest 7-day moving averages of deaths in the region, with Russia at 157.57 per 100,000 persons for the week of October 6 and Turkey at 60.43. Together, they accounted for approximately 90% of the deaths reported in the region. Up to October 13, the region had reported 38,461 cumulative deaths. For the week of 10/06, North Macedonia and Armenia had the highest death rates per 100,000 persons in the region, at 0.38 and 0.20, respectively. Armenia maintained a death rate of 0.20 per 100,000 persons the following week. Georgia had the highest death rate the week of 10/13 at 0.24 per 100,000 persons, up from 0.11 the previous week.

The 1-day persistence is an indicator of a clustering effect where an event on a particular day causes an increase in the number of cases on adjoining days. As shown in Table 6, the 1-day persistence was highest in Armenia at 16.1, and it increased to 21.5 the following week. North Macedonia saw a large increase from 8.8 to 16.5, with Georgia and Russia seeing smaller increases from 13.0 to 14.8 and 7.0 to 9.1, respectively.

Largely consistent with infection rates, Armenia, Georgia, and North Macedonia had the highest speed or average of new daily cases per 100,000 persons. During the week of October 6, Armenia had a speed of 15.4, increasing to 21.7 the following week. Georgia had a speed of 13.0, which increased slightly the following week to 13.8. North Macedonia had a speed of 9.0, which increased to 14.4. The region overall had an increase in speed from 4.2 to 5.2.

Speed is best used in conjunction with acceleration and jerk, which can provide further insight into potential pandemic trajectory changes. Four countries in the region had significant positive accelerations for the week of October 6: Georgia at 0.9, North Macedonia at 0.8, Armenia at 0.4, and Russia at 0.3. North Macedonia, Armenia, and Russia also had positive jerks. During the following week, in addition to the highest speed, Armenia had the highest acceleration and jerk in the region. Figure 2 shows shows weekly trends of SARS-CoV-2 in Central Asia [86], and additional trends can be found in Multimedia Appendices 1-5.

Figure 2. Weekly SARS-CoV2 trends in Central Asia [86].



Discussion

Principal Results

COVID-19 poses a significant threat to the Central Asian region, which is largely composed of former Soviet republics. These countries continue to suffer from food insecurity, high levels of poverty, and variation in health care quality and access as the region continues on its journey of transitioning from a centralized Soviet medical system. The population also suffers from multiple endemic infections, such as HIV/AIDS and tuberculosis. Russia and Turkey comprise the bulk of the population in Central Asia, and these countries are facing growing burdens of chronic disease and some of the highest obesity rates in Europe. Due to the combination of these factors, the region is vulnerable to negative outcomes from the COVID-19 pandemic. To date, the region has seen variations in policy intervention to control the spread of COVID-19 and mitigate outbreaks. Some countries, such as Kazakhstan and Azerbaijan, imposed strict and early lockdowns, while others, such as Russia, imposed more limited interventions.

Metrics tracking the progression of COVID-19 in Central Asia to date have largely been static, including measures such as new cases, cumulative cases, deaths, and 7-day moving averages. These metrics provide a view of the current state of the pandemic but are unable to provide any insight into the change in the speed of the pandemic over time or potential shifts in its trajectory, evolving from controlled spread to rapid growth or vice versa. These metrics also provide limited utility in comparing countries to each other and in analyzing countries with smaller populations. Novel metrics such as speed, acceleration, and jerk help contextualize static metrics and provide a view of trajectory over time, enabling potential anticipation of how the pandemic will evolve in the future.

Considering the static and dynamic metrics, it is apparent that the Central Asian region is trending negatively. The region saw an increase in 7-day moving average of new cases, infection rate, and speed for the week of October 13 compared to the week of October 6, with continued positive acceleration. This trend is largely driven by Russia and Turkey, which together encompass over 70% of the region's population and showed the highest 7-day moving averages in the region. Russia has the fourth highest speed in the region and continues to experience positive acceleration.

Kazakhstan, the fourth most populous country in the region, had the lowest infection rate for the week of October 6. This is

likely due to a continued emphasis on policy interventions to curb the spread of COVID-19. Authorities in Kazakhstan took some of the earliest precautions to prevent infections and continue to strictly enforce pandemic mitigation measures, including halting the easing of restrictions due to the global spike in COVID-19 cases in recent weeks.

Turkey, while contributing a significant portion of the total cases in the region due to its population size, has maintained relatively low infection rates of 1.81 and 1.96 per 100,000 persons for the weeks of October 6 and 13, respectively. Turkey has taken significant precautions to mitigate the spread of the virus, and authorities continue to enforce rigorous mask wearing and social distancing guidelines along with local quarantines when necessary.

Armenia is experiencing uncontrolled spread of COVID-19, with an infection rate of 13.73 per 100,000 persons for the week of October 6 quickly jumping to 25.19, the highest in the region, the following week. The pandemic speed, consistent with the infection rate trajectory, increased from 15.4 to 21.7, with an acceleration increase from 0.4 to 1.6. This change is likely due to the recent lifting of the COVID-19 state of emergency, which allowed the resumption of in-person schooling and international flights, among other activities.

After Armenia, Georgia and Russia had the highest infection rates in the region for the week of October 13. Russia continues to resist implementing interventions to curb the spread of COVID-19, with no mask mandates, capacity caps, or nightlife restrictions. In addition to the significant focus on developing an effective vaccine, there has been limited intervention to manage the spread of COVID-19 in Russia. This policy stance is impacting the trajectory of the region, which is trending negatively with no sign of a reversal in sight.

Conclusion

The rapid evolution, novel outbreaks, and frequently fluctuating trajectory of COVID-19 cannot be adequately assessed using static public health measures alone. Static measures, including the number of new COVID-19 cases, number of cumulative COVID-19 cases, 7-day moving average of new cases, rate of infection, number of new deaths, number of cumulative deaths, 7-day moving average of number of deaths, and death rates, provide a current view of the state of the pandemic. However, these measures do not provide any insight into how the trajectory of the pandemic may change over time.

Generally, the approach to modeling the spread of COVID-19 is to assume there is an underlying contagion model [87] and then to attempt to measure those model parameters [88], which involves contact tracing to determine the spread of the disease [89-92]. With an incubation period of up to 14 days [93], modeling this spread can take months. Early estimates of COVID-19 were developed using the method by Lipsitch et al [94], which was used for contact tracing in Wuhan and Italy; however, the statistical properties were weak [95-100]. Zhao et al [96] estimated the serial interval distribution and R_0 based on only 6 pairs of cases, which is insufficient to understand the transmission of COVID-19 [101]. This results in relaxing of the assumptions in these models, such as disaggregating the population by geography and modeling within-geography and across-geography personal interactions [102]. Martcheva [103] developed a dynamic model from several contagion models and their possible dynamics [104,105]. They are limited to the statistical inference of parameter values from actual data [106].

Novel surveillance metrics allow for a more nuanced analysis of the COVID-19 pandemic and together with static metrics, can enable policymakers to make informed decisions to control the spread of the pandemic and prevent further outbreaks. Novel dynamic metrics include speed, acceleration, jerk, and 7-day persistence, and they provide potential insight into how the pandemic will evolve in the future.

The analysis of Central Asia using static and novel surveillance metrics suggests that the region is precariously positioned and trending negatively. Russia, the largest country by population, continues to have high infection rates and one of the highest speeds of infection. With no sign of increasing restrictions, it is unlikely that this trend will reverse and that outbreaks in the region will be controlled.

Limitations

Our data are limited by reporting methods across individual countries. Some countries, such as Turkmenistan, refuse to acknowledge COVID-19 cases. Variation in testing and infrastructure may impact the number of cases reported by other countries. The data are reported at a national level, which does not enable any subnational analysis.

Comparison to Prior Work

This study is part of a broader research program at Northwestern Feinberg School of Medicine (The Global SARS-CoV-2 Surveillance Project: Policy, Persistence, & Transmission). Novel surveillance metrics, including speed, acceleration, jerk, and 7-day persistence, have been developed by this research program and are being applied to all global regions.

Acknowledgments

This publication was made possible through support provided by Feed the Future through the US Agency for International Development under the terms of Contract No. 7200LA1800003. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the US Agency for International Development.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Weekly Cental Asia SARS-CoV-2 statistics by country.
[PNG File , 154 KB - jmir_v23i2e25799_app1.png]

Multimedia Appendix 2
Weekly Central Asia 7-day persistence map.
[PNG File , 372 KB - jmir_v23i2e25799_app2.png]

Multimedia Appendix 3
Weekly Central Asia statistics.
[PNG File , 199 KB - jmir_v23i2e25799_app3.png]

Multimedia Appendix 4
Weekly Central Asia jerk map.
[PNG File , 359 KB - jmir_v23i2e25799_app4.png]

Multimedia Appendix 5
Weekly Central Asia acceleration jerk map.
[PNG File , 434 KB - jmir_v23i2e25799_app5.png]

References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
2. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-12-09]
3. McKee M, Stuckler D. If the world fails to protect the economy, COVID-19 will damage health not just now but also in the future. Nature Medicine. 2020 ;26(5):640-642.
4. The World Bank in Europe and Central Asia. The World Bank. URL: https://www.worldbank.org/en/region/eca [accessed 2020-08-10]
5. Collapse of the Soviet Union. Encyclopaedia Britannica. URL: https://www.britannica.com/event/the-collapse-of-the-Soviet-Union#ref1258663 [accessed 2021-01-25]
6. D'Anieri P. Explaining the success and failure of post-communist revolutions. CPCS 2006;39(3):331-350. [doi: 10.1016/j.postcomstud.2006.06.002]
7. Adambekov S, Kaiyrlykyzy A, Igissinov N, Linkov F. Health challenges in Kazakhstan and Central Asia. J Epidemiol Community Health 2016;70(1):104-108. [Medline: 26254293]
8. Schluger NW, El-Bassel N, Hermosilla S, Terlikbayeva A, Darisheva M, Aifah A, et al. Tuberculosis, drug use and HIV infection in Central Asia: an urgent need for attention. Drug Alcohol Depend 2013 Nov;132 Suppl 1:S32-S36. [doi: 10.1016/j.drugalcdep.2013.07.012] [Medline: 23928052]
9. Altice FL, Azbel L, Stone J, Brooks-Pollock E, Smyrnov P, Dvoriak S, et al. The perfect storm: incarceration and the high-risk environment perpetuating transmission of HIV, hepatitis C virus, and tuberculosis in Eastern Europe and Central Asia. Lancet 2016 Sep;388(10050):1228-1248. [doi: 10.1016/s0140-6736(16)30856-x]
10. Stickley A, Leinsalu M, Andreev E, Razvodovsky Y, Vågerö D, McKee M. Alcohol poisoning in Russia and the countries in the European part of the former Soviet Union, 1970 2002. Eur J Public Health 2007 Oct 27;17(5):444-449. [doi: 10.1093/eurpub/ckl275] [Medline: 17327281]
11. Leon DA, Chenet L, Shkolnikov VM, Zakharov S, Shapiro J, Rakhmanova G, et al. Huge variation in Russian mortality rates 1984–94: artefact, alcohol, or what? Lancet 1997 Aug;350(9075):383-388. [doi: 10.1016/s0140-6736(97)03360-6]
12. Nemtsov A. Alcohol-related human losses in Russia in the 1980s and 1990s. Addiction 2002 Nov;97(11):1413-1425. [doi: 10.1046/j.1360-0443.2002.00262.x] [Medline: 12410782]
13. Wadhwa D. Men smoke 5 times more than women. World Bank Blogs. 2019 May 23. URL: https://blogs.worldbank.org/opendata/men-smoke-5-times-more-women#:~:text=Overall%2C%20smoking%20is%20most%20prevalent,by%20gender%20is%20more%20extreme [accessed 2020-08-11]
14. Stockwell T, Andreasson S, Cherpitel C, Chikritzhs T, Dangardt F, Holder H, et al. The burden of alcohol on health care during COVID-19. Drug Alcohol Rev 2021 Jan;40(1):3-7 [FREE Full text] [doi: 10.1111/dar.13143] [Medline: 32835427]
15. Saengow U, Assanangkornchai S, Casswell S. Alcohol: a probable risk factor of COVID-19 severity. Addiction 2021 Jan 12;116(1):204-205 [FREE Full text] [doi: 10.1111/add.15194] [Medline: 32688440]

16. Spagnolo PA, Montemitro C, Leggio L. New Challenges in Addiction Medicine: COVID-19 Infection in Patients With Alcohol and Substance Use Disorders-The Perfect Storm. Am J Psychiatry 2020 Sep 01;177(9):805-807. [doi: 10.1176/appi.ajp.2020.20040417] [Medline: 32660296]
17. Hamer M, Kivimäki M, Gale CR, Batty GD. Lifestyle risk factors, inflammatory mechanisms, and COVID-19 hospitalization: A community-based cohort study of 387,109 adults in UK. Brain Behav Immun 2020 Jul;87:184-187 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.059] [Medline: 32454138]
18. Volkow N. Collision of the COVID-19 and Addiction Epidemics. Ann Intern Med 2020 Jul 07;173(1):61-62 [FREE Full text] [doi: 10.7326/M20-1212] [Medline: 32240293]
19. Hochstatter KR, Akhtar WZ, Dietz S, Pe-Romashko K, Gustafson DH, Shah DV, et al. Potential Influences of the COVID-19 Pandemic on Drug Use and HIV Care Among People Living with HIV and Substance Use Disorders: Experience from a Pilot mHealth Intervention. AIDS Behav 2020 Jul 23 [FREE Full text] [doi: 10.1007/s10461-020-02976-1] [Medline: 32705370]
20. Biddle N, Edwards B, Gray M, Sollis K. Alcohol consumption during the COVID-19 period: May 2020. ANU Centre for Social Research and Methods. 2020 May. URL: https://csrm.cass.anu.edu.au/sites/default/files/docs/2020/6/Alcohol_consumption_during_the_COVID-19_period.pdf [accessed 2021-01-25]
21. Frishman WH, Del Vecchio A, Sanal S, Ismail A. Cardiovascular manifestations of substance abuse: part 2: alcohol, amphetamines, heroin, cannabis, and caffeine. Heart Dis 2003;5(4):253-271. [doi: 10.1097/01.hdx.0000080713.09303.a6] [Medline: 12877759]
22. Wu L, Zhu H, Ghitza UE. Multicomorbidity of chronic diseases and substance use disorders and their association with hospitalization: Results from electronic health records data. Drug Alcohol Depend 2018 Nov 01;192:316-323 [FREE Full text] [doi: 10.1016/j.drugalcdep.2018.08.013] [Medline: 30312893]
23. Degenhardt L, Charlson F, Ferrari A, Santomauro D, Erskine H, Mantilla-Herrara A, et al. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Psychiatry 2018 Dec;5(12):987-1012. [doi: 10.1016/s2215-0366(18)30337-7]
24. COVID-19: Avoiding a 'second tragedy' for those with TB. Médecins Sans Frontières. 2020 Mar 23. URL: https://www.msf.org/covid-19-how-avoid-second-tragedy-those-tuberculosis [accessed 2020-08-09]
25. Fossat M. In remote Kyrgyzstan, COVID-19 puts a strain on the health system. Médecins Sans Frontières. 2020 Jul 22. URL: https://www.msf.org/kyrgyzstan-covid-19-strains-health-system [accessed 2020-08-09]
26. Sedik D. Europe and Central Asia - Regional overview of food insecurity. Food and Agriculture Organizations of the United Nations. 2017. URL: http://www.fao.org/3/a-i6877e.pdf [accessed 2021-01-25]
27. Europe and Central Asia Regional Overview of Food Security and Nutrition. Food and Agriculture Organization of the United Nations. 2019. URL: http://www.fao.org/3/ca7153en/CA7153EN.pdf [accessed 2021-01-25]
28. The State of Food Security and Nutrition in the World 2019. Food and Agriculture Organization of the United Nations. 2019. URL: https://docs.wfp.org/api/documents/WFP-0000106760/download/?_ga=2.37517920.1134316896.1611609114-2001096053.1611609114 [accessed 2021-01-25]
29. William SA. Economic and Social Impacts of COVID-19: Updates from the Listening to Tajikistan Survey. World Bank. 2020 Jul 13. URL: https://www.worldbank.org/en/news/factsheet/2020/07/13/economic-and-social-impacts-of-covid-19-update-from-listening-to-tajikistan [accessed 2020-08-05]
30. Sapkota A. Water reuse, food production and public health: Adopting transdisciplinary, systems-based approaches to achieve water and food security in a changing climate. Environ Res 2019 Apr;171:576-580. [doi: 10.1016/j.envres.2018.11.003] [Medline: 30477872]
31. Micklin P. The Aral Sea disaster. Annu Rev Earth Planet Sci 2007 May;35(1):47-72. [doi: 10.1146/annurev.earth.35.031306.140120]
32. Burunciuc L. Improving water and sanitation in Central Asia requires determination and shared commitment. World Bank Blogs. 2019 Dec 24. URL: https://blogs.worldbank.org/europeandcentralasia/improving-water-and-sanitation-central-asia-requires-determination-and-shared [accessed 2020-08-03]
33. Ataniyazova OA. COVID-19: High risk groups for COVID-19 and challenges for medical universities. Asian Pac J Envi & Cancer 2020 Jul 17;3(S1):21-23. [doi: 10.31557/apjec.2020.3.s1.21-23]
34. Azam R, Idrees A. Economic growth, obesity and global burden of disease: an exploration. GCU Economic Journal 2019;LII(1&2):67-106.
35. Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Martínez-Arias R, et al. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. Am J Hum Genet 1998 Dec;63(6):1824-1838 [FREE Full text] [doi: 10.1086/302133] [Medline: 9837835]
36. Kuzmina E. In: Mair VH, editor. The Prehistory of the Silk Road. Philadelphia, PA: University of Pennsylvania Press; 2008.
37. Powles J, Fahimi S, Micha R, Khatibzadeh S, Shi P, Ezzati M, Global Burden of Diseases Nutrition and Chronic Diseases Expert Group (NutriCoDE). Global, regional and national sodium intakes in 1990 and 2010: a systematic analysis of 24 h urinary sodium excretion and dietary surveys worldwide. BMJ Open 2013 Dec 23;3(12):e003733 [FREE Full text] [doi: 10.1136/bmjopen-2013-003733] [Medline: 24366578]

38. Krause M, Gerchman F, Friedman R. Coronavirus infection (SARS-CoV-2) in obesity and diabetes comorbidities: is heat shock response determinant for the disease complications? Diabetol Metab Syndr 2020 Jul 16;12(1):63 [FREE Full text] [doi: 10.1186/s13098-020-00572-w] [Medline: 32690985]
39. Maukayeva S, Karimova S. Epidemiologic character of COVID-19 in Kazakhstan: A preliminary report. North Clin Istanb 2020;7(3):210-213 [FREE Full text] [doi: 10.14744/nci.2020.62443] [Medline: 32478290]
40. Global Economic Prospects: Europe and Central Asia. World Bank. 2020 Jun. URL: https://www.worldbank.org/en/region/eca/brief/global-economic-prospects-europe-and-central-asia [accessed 2020-08-05]
41. Azamat AI, Izvorski IV, Rahardja S. Kazakhstan Economic Update: Navigating the Crisis. World Bank. 2020 Jul 22. URL: https://www.worldbank.org/en/country/kazakhstan/publication/economic-update-summer-2020 [accessed 2020-08-04]
42. Sumner A, Hoy C, Ortiz-Juarez E. Estimates of the Impact of COVID-19 on Global Poverty. UNU-WIDER. 2020 Apr. URL: https://www.wider.unu.edu/sites/default/files/Publications/Working-paper/PDF/wp2020-43.pdf [accessed 2021-01-25]
43. Valensisi G. COVID-19 and global poverty. Food and Agriculture Organization of the United Nations. 2020. URL: http://www.fao.org/3/ca8824en/CA8824EN.pdf [accessed 2021-01-25]
44. Nogales R, Oldiges C, Dirksen J, Alkire S. Multidimensional Poverty and COVID-19 Risk Factors: A Rapid Overview of Interlinked Deprivations across 5.7 Billion People. Oxford Poverty & Human Development Initiative. 2020. URL: https://ophi.org.uk/wp-content/uploads/B53_Covid-19_vs3-2_2020_online.pdf [accessed 2021-01-25]
45. Zar HJ, Dawa J, Fischer GB, Castro-Rodriguez JA. Challenges of COVID-19 in children in low- and middle-income countries. Paediatr Respir Rev 2020 Sep;35:70-74 [FREE Full text] [doi: 10.1016/j.prrv.2020.06.016] [Medline: 32654854]
46. Dutta P, Andrews C, de MA, Clay T, Chaudhary S. Economic Inclusion for the Poorest and COVID-19: Adaptation and Early Priorities for Medium- and Longer-Term Recovery. World Bank. 2020 Jun 18. URL: https://reliefweb.int/sites/reliefweb.int/files/resources/77523_0.pdf [accessed 2021-01-25]
47. Russia: Health Workers Face Retaliation for Speaking Out. Human Rights Watch. 2020 Jun 15. URL: https://www.hrw.org/news/2020/06/15/russia-health-workers-face-retaliation-speaking-out [accessed 2020-08-09]
48. Turkey: Probes Over Doctors' Covid-19 Comments. Human Rights Watch. 2020 Jun 10. URL: https://www.hrw.org/news/2020/06/10/turkey-probes-over-doctors-covid-19-comments [accessed 2020-08-09]
49. Turkmenistan Denies Apparent Covid-19 Outbreak. Human Rights Watch. 2020 Jun 07. URL: https://www.hrw.org/news/2020/06/27/turkmenistan-denies-apparent-covid-19-outbreak [accessed 2020-08-09]
50. Huseynov V. Azerbaijan hopes to minimize ramifications of coronavirus crisis?Analysis. The Jamestown Foundation. 2020 Apr 16. URL: https://jamestown.org/program/azerbaijan-hopes-to-minimize-ramifications-of-coronavirus-crisis/ [accessed 2021-01-25]
51. Samadov B. Azerbaijan—COVID-19 and a Divided Opposition. The COVID-19 Pandemic in the South Caucasus. 2020 May. URL: https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/415805/CAD115.pdf?sequence=2&isAllowed=y [accessed 2021-01-25]
52. Armenia: Law Restricts Privacy Amid COVID-19 Fight. Human Rights Watch. 2020 Apr 03. URL: https://www.hrw.org/news/2020/04/03/armenia-law-restricts-privacy-amid-covid-19-fight [accessed 2020-08-09]
53. Aitkhozhina D. Most Covid-19 restrictions lifted, but Moscow is not open for protest. Human Rights Watch. 2020 Jul 22. URL: https://www.hrw.org/news/2020/07/22/most-covid-19-restrictions-lifted-moscow-not-open-protest [accessed 2020-08-09]
54. IFPRI's COVID-19 policy response (CPR) portal: Identifying trends and implications for food systems. Nigeria Strategy Support Program. 2020 May 20. URL: http://nssp.ifpri.info/2020/05/20/ifpris-covid-19-policy-response-cpr-portal-identifying-trends-and-implications-for-food-systems/ [accessed 2021-01-25]
55. Azerbaijan Monthly Economic Update. World Bank Group. 2020 Jul. URL: http://pubdocs.worldbank.org/en/443481594846767557/AZ-Monthly-Economic-Update-July-2020.pdf [accessed 2020-08-04]
56. Callaway E. Russia's fast-track coronavirus vaccine draws outrage over safety. Nature 2020 Aug;584(7821):334-335. [doi: 10.1038/d41586-020-02386-2] [Medline: 32782400]
57. Zimmer C. 'This Is All Beyond Stupid.' Experts Worry About Russia's Rushed Vaccine. New York Times. 2020 Aug 11. URL: https://www.nytimes.com/2020/08/11/health/russia-covid-19-vaccine-safety.html [accessed 2020-08-14]
58. Neuman S. Putin Announces Approval Of A 2nd Unproven Coronavirus Vaccine. NPR. 2020 Oct 14. URL: https://www.npr.org/sections/coronavirus-live-updates/2020/10/14/923785152/russias-putin-announces-approval-of-a-second-unproven-coronavirus-vaccine [accessed 2020-01-25]
59. Sherwin E. Russia's 'corona-holidays' leave small businesses in freefall. DW. 2020 Apr 03. URL: https://p.dw.com/p/3aQXo [accessed 2020-10-19]
60. WorldOMeter. Russia. Worldometer. URL: https://www.worldometers.info/coronavirus/country/russia/ [accessed 2020-10-19]
61. Sherwin E. Coronavirus: Russia tightens restrictions, but avoids lockdown. DW. 2020 Oct 16. URL: https://p.dw.com/p/3k2V8 [accessed 2021-01-25]
62. Coronavirus in Russia: The Latest News. The Moscow Times. URL: https://www.themoscowtimes.com/2020/10/19/coronavirus-in-russia-the-latest-news-oct-19-a69117 [accessed 2020-10-19]
63. Teutsch SM. Considerations in Planning a Surveillance System. In: Principles & Practice of Public Health Surveillance. Oxford, UK: Oxford Scholarship Online; 2010:18-28.

64. Teutsch S, Churchill R. Principles and Practice of Public Health Surveillance. Oxford, UK: Oxford University Press; 2000.
65. Teutsch SM, Thacker SB. Planning a public health surveillance system. Epidemiological bulletin 1995;16(1):1.
66. Thacker S, Berkelman R. Public health surveillance in the United States. Epidemiol Rev 1988;10(1):164-190. [doi: 10.1093/oxfordjournals.epirev.a036021] [Medline: 3066626]
67. Thacker S, Choi K, Brachman P. The surveillance of infectious diseases. JAMA 1983 Mar 04;249(9):1181-1185. [Medline: 6823080]
68. Thacker S, Stroup D. Future directions for comprehensive public health surveillance and health information systems in the United States. Am J Epidemiol 1994 Sep 01;140(5):383-397. [doi: 10.1093/oxfordjournals.aje.a117261] [Medline: 8067331]
69. Thacker S, Stroup D. Public Health Surveillance and Health Services Research. In: Armenian HK, Shapiro S, editors. Epidemiology and Health Services. New York, NY: Oxford University Press; 1998:61-82.
70. Thacker SB, Stroup DF, Parrish RG, Anderson HA. Surveillance in environmental public health: issues, systems, and sources. Am J Public Health 1996 May;86(5):633-638. [doi: 10.2105/ajph.86.5.633] [Medline: 8629712]
71. Coronavirus Disease 2019 (COVID-19). US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/index.html [accessed 2021-01-25]
72. Global Surveillance for human infection with coronavirus disease (COVID-19). World Health Organization. URL: https://apps.who.int/iris/handle/10665/330857 [accessed 2021-01-25]
73. EU level surveillance of COVID-19. European Center for Disease Prevention and Control. URL: https://www.ecdc.europa.eu/en/covid-19/surveillance [accessed 2021-01-25]
74. COVID-19 Surveillance Dashboard. African Union. URL: https://au.int/en/covid19 [accessed 2021-01-25]
75. Hannah RE, Diana B, Edouard M, Joe H, Bobbie M, Charlie G, et al. Coronavirus Pandemic (COVID-19) – the data. Our World in Data. URL: https://ourworldindata.org/coronavirus-data [accessed 2020-08-14]
76. COVID-19 Situation Dashboard. European Centre for Disease Prevention and Control. URL: https://qap.ecdc.europa.eu/public/extensions/COVID-19/COVID-19.html [accessed 2021-01-25]
77. Coronavirus Resource Center. Johns Hopkins University of Medicine. URL: https://coronavirus.jhu.edu/ [accessed 2021-01-25]
78. Coronavirus Outbreak. Bloomberg. URL: https://www.bloomberg.com/coronavirus [accessed 2021-01-25]
79. US Coronavirus Cases and Deaths: Track COVID-19 data daily by state and county. USA Facts. URL: https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/ [accessed 2021-01-25]
80. Oehmke J, Moss C, Singh L, Oehmke T, Post L. Dynamic Panel Surveillance of COVID-19 Transmission in the United States to Inform Health Policy: Observational Statistical Study. J Med Internet Res 2020 Oct 05;22(10):e21955 [FREE Full text] [doi: 10.2196/21955] [Medline: 32924962]
81. Oehmke J, Oehmke T, Singh L, Post L. Dynamic Panel Estimate-Based Health Surveillance of SARS-CoV-2 Infection Rates to Inform Public Health Policy: Model Development and Validation. J Med Internet Res 2020 Sep 22;22(9):e20924 [FREE Full text] [doi: 10.2196/20924] [Medline: 32915762]
82. Abdurasulov A. Coronavirus: Why has Turkmenistan reported no cases? BBC. 2020 Apr 07. URL: https://www.bbc.com/news/world-asia-52186521 [accessed 2020-08-14]
83. SARS-CoV-2 Test Tracker. Foundation for Innovative New Diagnostics. URL: https://www.finddx.org/covid-19/test-tracker/ [accessed 2020-10-14]
84. FINDCov19TrackerData. GitHub. URL: https://github.com/dsbbfinddx/FINDCov19TrackerData [accessed 2021-01-25]
85. Hansen LP. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 1982 Jul;50(4):1029. [doi: 10.2307/1912775]
86. Post L. Global SARS-CoV-2 Surveillance Policy, Persistence, and Transmission - Central Asia. Tableau Public. 2020 Oct 22. URL: https://public.tableau.com/profile/lori.post#!/vizhome/GlobalSARS-CoV-2SurveillancePolicyPersistenceandTransmission-CentralAsia/Story [accessed 2021-01-26]
87. Iacopini I, Petri G, Barrat A, Latora V. Simplicial models of social contagion. Nat Commun 2019 Jun 06;10(1):2485 [FREE Full text] [doi: 10.1038/s41467-019-10431-6] [Medline: 31171784]
88. Stock JH. Data gaps and the policy response to the novel coronavirus. National Bureau of Economic Research. 2020. URL: https://www.nber.org/system/files/working_papers/w26902/w26902.pdf [accessed 2021-01-25]
89. Desjardins M, Hohl A, Delmelle E. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. Appl Geogr 2020 May;118:102202 [FREE Full text] [doi: 10.1016/j.apgeog.2020.102202] [Medline: 32287518]
90. Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, et al. Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19. IEEE Rev Biomed Eng 2021;14:4-15. [doi: 10.1109/rbme.2020.2987975]
91. Naudé W. Artificial Intelligence against COVID-19: An early review. IZA Institute of Labor Economics. 2020 Apr. URL: http://ftp.iza.org/dp13110.pdf [accessed 2021-01-25]
92. Tang Z, Zhao W, Xie X, Zhong Z, Shi F, Liu J, et al. Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images. ArXiv. Preprint posted online on March 26, 2020 [FREE Full text]
93. Interim clinical guidance for management of patients with confirmed coronavirus disease (COVID-19). Centers for Disease Control and Prevention. 2020 Jun 30. URL: https://stacks.cdc.gov/view/cdc/89980 [accessed 2021-01-25]

94. Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, et al. Transmission dynamics and control of severe acute respiratory syndrome. Science 2003 Jun 20;300(5627):1966-1970 [FREE Full text] [doi: 10.1126/science.1086616] [Medline: 12766207]
95. Lipsitch M, Swerdlow DL, Finelli L. Defining the epidemiology of COVID-19 — studies needed. N Engl J Med 2020 Mar 26;382(13):1194-1196. [doi: 10.1056/nejmp2002125]
96. Zhao S, Gao D, Zhuang Z, Chong MKC, Cai Y, Ran J, et al. Estimating the serial interval of the novel coronavirus disease (COVID-19): A statistical analysis using the public data in Hong Kong from January 16 to February 15, 2020. Research Square. Preprint posted online on May 27, 2020 [FREE Full text] [doi: 10.21203/rs.3.rs-18805/v3]
97. Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, et al. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. Proc Natl Acad Sci USA 2020 May 12;117(19):10484-10491 [FREE Full text] [doi: 10.1073/pnas.2004978117] [Medline: 32327608]
98. You C, Deng Y, Hu W, Sun J, Lin Q, Zhou F, et al. Estimation of the time-varying reproduction number of COVID-19 outbreak in China. Int J Hyg Environ Health 2020 Jul;228:113555 [FREE Full text] [doi: 10.1016/j.ijheh.2020.113555] [Medline: 32460229]
99. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. Emerg Infect Dis 2020 Jul;26(7):1470-1477 [FREE Full text] [doi: 10.3201/eid2607.200282] [Medline: 32255761]
100. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 2020 Jun 22;26(6):855-860 [FREE Full text] [doi: 10.1038/s41591-020-0883-7] [Medline: 32322102]
101. Nishiura H, Linton N, Akhmetzhanov A. Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis 2020 Apr;93:284-286 [FREE Full text] [doi: 10.1016/j.ijid.2020.02.060] [Medline: 32145466]
102. Lai C, Shih T, Ko W, Tang H, Hsueh P. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020 Mar;55(3):105924 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.105924] [Medline: 32081636]
103. Martcheva M, Castillo-Chavez C. Diseases with chronic stage in a population with varying size. Mathematical Biosciences 2003 Mar;182(1):1-25. [doi: 10.1016/s0025-5564(02)00184-0]
104. Chen L. Persistent spatial patterns of interacting contagions. Phys. Rev. E 2019 Feb 14;99(2):022308. [doi: 10.1103/physreve.99.022308]
105. Kwuimy CAK, Nazari F, Jiao X, Rohani P, Nataraj C. Nonlinear dynamic analysis of an epidemiological model for COVID-19 including public behavior and government action. Nonlinear Dyn 2020 Jul 16;101(3):1-15 [FREE Full text] [doi: 10.1007/s11071-020-05815-z] [Medline: 32836814]
106. Li X, Yang J, Martcheva M. Age Structured Epidemic Modeling. New York, NY: Springer; 2019.

Abbreviations

FIND: Foundation for Innovative New Diagnostics
R_0: basic reproduction number
USSR: Union of Soviet Socialist Republics

Edited by G Eysenbach; submitted 16.11.20; peer-reviewed by A Seidmann; comments to author 07.12.20; revised version received 09.12.20; accepted 20.01.21; published 03.02.21.

Please cite as:
Post LA, Benishay ET, Moss CB, Murphy RL, Achenbach CJ, Ison MG, Resnick D, Singh LN, White J, Chaudhury AS, Boctor MJ, Welch SB, Oehmke JF
Surveillance Metrics of SARS-CoV-2 Transmission in Central Asia: Longitudinal Trend Analysis
J Med Internet Res 2021;23(2):e25799
URL: https://www.jmir.org/2021/2/e25799
doi:10.2196/25799
PMID:33475513

Original Paper

Using Twitter to Understand the COVID-19 Experiences of People With Dementia: Infodemiology Study

Juanita-Dawne Bacsu[1], PhD; Megan E O'Connell[1], PhD; Allison Cammer[2], PhD; Mahsa Azizi[3], MSc; Karl Grewal[1], BSc; Lisa Poole[4], BA; Shoshana Green[1], MEd; Saskia Sivananthan[5], PhD; Raymond J Spiteri[3], PhD

[1]Department of Psychology, University of Saskatchewan, Saskatoon, SK, Canada

[2]College of Pharmacy and Nutrition, University of Saskatchewan, Saskatoon, SK, Canada

[3]Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada

[4]Dementia Advocacy Canada, Calgary, AB, Canada

[5]Alzheimer Society of Canada, Toronto, ON, Canada

Corresponding Author:
Megan E O'Connell, PhD
Department of Psychology
University of Saskatchewan
9 Campus Drive
Saskatoon, SK, S7N 5A5
Canada
Phone: 1 306 966 2496
Email: megan.oconnell@usask.ca

Abstract

Background: The COVID-19 pandemic is affecting people with dementia in numerous ways. Nevertheless, there is a paucity of research on the COVID-19 impact on people with dementia and their care partners.

Objective: Using Twitter, the purpose of this study is to understand the experiences of COVID-19 for people with dementia and their care partners.

Methods: We collected tweets on COVID-19 and dementia using the GetOldTweets application in Python from February 15 to September 7, 2020. Thematic analysis was used to analyze the tweets.

Results: From the 5063 tweets analyzed with line-by-line coding, we identified 4 main themes including (1) separation and loss; (2) COVID-19 confusion, despair, and abandonment; (3) stress and exhaustion exacerbation; and (4) unpaid sacrifices by formal care providers.

Conclusions: There is an imminent need for governments to rethink using a one-size-fits-all response to COVID-19 policy and use a collaborative approach to support people with dementia. Collaboration and more evidence-informed research are essential to reducing COVID-19 mortality and improving the quality of life for people with dementia and their care partners.

(J Med Internet Res 2021;23(2):e26254) doi:10.2196/26254

KEYWORDS

Twitter; social media; dementia; COVID-19; health policy; experience; support; disorder; theme; collaborate; quality of life

Introduction

The COVID-19 pandemic is having an impact on people with dementia. In Canada, almost two-thirds of all COVID-19–related deaths have been people with dementia [1]. In the United Kingdom, 50% of COVID-19–related deaths in care homes have been people with dementia [2].

In comparison to other groups, people with dementia are among the most vulnerable to the COVID-19 pandemic [3]. Although people with dementia are not necessarily more susceptible to COVID-19, advancing age, frailty, and coexisting health conditions often associated with dementia (eg, cardiovascular disease, hypertension, or diabetes) increase the risk of complications [4,5]. Moreover, social isolation measures (eg, stay-at-home orders, visitation bans, and lockdowns in care facilities) from COVID-19 may increase the risk of hospitalization and mortality for people with dementia [6].

People with dementia also face a heightened risk of COVID-19 exposure due to cognitive impairment and memory loss [3]. For example, people with dementia may have challenges remembering or understanding self-protection protocols such as wearing a mask, using proper hand hygiene, and maintaining physical distance from others [7,8]. Despite these challenges, there is a paucity of knowledge on the COVID-19 pandemic's impact on people with dementia.

There is an urgent need to understand the experiences of people with dementia and their care partners in the COVID-19 pandemic [8-10]. Mitigation strategies that include physical distancing make it difficult to conduct timely and collaborative research, with many universities suspending recruitment and in-person studies during the pandemic [11]. Given these difficulties, examining the impact of COVID-19 requires ingenuity and innovation.

With over 330 million monthly users [12], the microblogging and social networking website Twitter [13] presents an innovative opportunity to examine the COVID-19 impact on people with dementia. For example, Twitter users are publicly sharing lived experiences of COVID-19 and dementia. The purpose of this study is to use Twitter to understand the COVID-19 experiences of people with dementia and their care partners.

Methods

Recruitment

We scraped tweets posted in the English language containing synonyms for dementia (eg, "Alzheimer's disease" and "Lewy Body disease") and synonyms for COVID-19 (including nontechnical phrases such as "Corona") during the period from February 15 to September 7, 2020, from Twitter using the GetOldTweets application programming interface in Python [14]. The Tweets were not geo-tagged.

Twitter is a social media and microblogging website where users share their posts with the public. Following on existing studies, there is a general consensus that tweets posted publicly on Twitter can be used for research [15,16]. Tweets on the Twitter website are located within the public domain; informed consent was not obtained.

Data Exclusion

From the approximately 20,800 tweets that were gathered using these criteria, we applied filters referring to candidates for the US presidential election (eg, synonyms for "Donald Trump" and "Joe Biden"). In addition, the filter "Tom Seaver" was used to delete tweets of public reactions to the major league baseball player who was reported to have died on August 31, 2020, of complications due to COVID-19 and dementia with Lewy bodies. Retweets and nonoriginal tweets were also excluded. Finally, to increase the likelihood of scraping tweets that described personal experiences with dementia during the era of COVID-19, we excluded tweets that did not include synonyms for familial relationships (eg, "father," "mother," and "grandparent") or friends and acquaintances (eg, "buddy," "pal," and "neighbour"). This filtering procedure resulted in a total of 5063 tweets that were extracted into an Excel (Microsoft Corporation) spreadsheet for qualitative line-by-line coding as the basis for thematic analysis.

Data Analysis: Coding and Intercoder Consistency

To develop a robust codebook, authors JDB and MEO read and reread 300 tweets. The researchers independently assigned codes to each of the 300 tweets. After coding the tweets, the two researchers met to discuss their code lists and developed an initial codebook. The initial codebook consisted of 18 codes with code definitions, keywords, and specific examples (eg, tweets).

To test intercoder consistency, a team of six researchers pilot-tested 100 tweets by independently coding the tweets according to the codebook. Once the team finished their coding, the codes were compared with a coding example sheet that JDB and MEO developed. Following this pilot test, a group coding workshop was held to collectively pilot-code 50 tweets and discuss any inconsistencies in the interpretation of the codes. During this workshop, group discussion resolved coding uncertainties and disputes, leading to the refinement of the codebook (eg, overlapping codes or unused codes). The final version of the codebook consisted of 9 codes including (1) death, (2) fear for person with dementia's health and well-being, (3) challenges and unmet needs, (4) separation or restricted visiting, (5) formal care provider challenges, (6) supports described, (7) informal caregiver's health and well-being, (8) stories of survival, and (9) user identifies as a person with dementia (used as a second tag vs a primary code).

The tweets (n=5063) were then divided among the seven authors (approximately 723 tweets each) for descriptive analysis, with JDB managing intercoder consistency throughout the coding process. For 7 consecutive days, each coder completed approximately 100 codes per day and sent them to JDB. Coders flagged any tweets of which they were unsure or uncertain for JDB to review. In addition to reviewing the flagged tweets, the JDB randomly reviewed 10% of each coder's work and provided feedback (eg, flagged tweets or any inconsistencies) for each day of coding. Throughout the coding process, JDB worked in direct collaboration with the MEO to discuss any discrepancies or uncertainties within the data. Once the coding was completed, two team meetings were held to identify and discuss the key themes arising from the data.

Results

From the 5063 tweets analyzed, we identified 4 main themes including (1) separation and loss; (2) COVID-19 confusion, despair, and abandonment; (3) stress and exhaustion; and (4) risks of exposure and personal sacrifices faced by formal care providers. Illustrative tweets are provided and are unedited for grammar.

Separation and Loss

This theme captures the psychological sense of loss due to physical barriers that create separation. Many personal descriptions included discussion of the physical separation, which includes separation due to death, separation during the process of dying during the COVID-19 pandemic, and separation due to visitation bans because of COVID-19. Underlying these

descriptors of physical separation is a clear psychological disconnection accompanied by feelings of loss. Separation creates a psychological sense of loss. A notable subtheme in separation and loss is the expression of the sense of loss expressed by care partners at the acceleration of decline perceived in persons with dementia during COVID-19, which was frequently blamed on the visitation bans and imposed lockdowns in care homes and health facilities. The intersection of the multiple dimensions of loss and separation are highlighted in the following illustrative quotes:

> *Yesterday I lost my mom. Due to covid [sic] I was unable to see her for the last few months. I did get to FaceTime twice and we did chat on the phone until her dementia made it difficult for her to do so. There will be no memorial until it is cleared to do so. My heart is broken.*

> *...My husband passed away a few days ago, victim of Covid [sic] protocol! He had dementia, didn't understand why I couldn't visit him. He lost hope, 36 lbs in 23 days; could not be saved. This is so cruel to do to our seniors/he was a veteran!!! WRONG!!!*

> *...Dad had dementia and was otherwise healthy. He had daily visitors even though he had little memory of who they were. Covid [sic] closed the nursing home to visitors and he stopped eating. He lost 40lbs and died in June. It's another aspect of the Covid [sic]. Dying Alone.*

> *Let me tell you what this covid [sic] lockdown did, it killed my daddy. He had dementia and he was still doing good, then the lockdown, we weren't there to hold him and to help feed him. When we went to see him, he was a shell, there was nothing left of him. I am so angry.*

COVID-19 Confusion, Despair, and Abandonment

Another predominant theme was feelings of despair and abandonment among people with dementia from COVID-19 confinement and visitation bans. Many tweets described difficulties understanding COVID-19 displayed or experienced by persons with dementia and negative psychological consequences due to this confusion about COVID-19. Many reported that people with dementia could not understand the changes necessitated by the pandemic response; they required constant teaching, reminders, and reassurance. For some, this lack of understanding of COVID-19 led to challenges (described or implied) of living with the new policies for social distancing, mask wearing, and other sanitary precautions, which engendered feelings of hopelessness and despair. For many others, the lack of understanding the physical distancing restrictions due to COVID-19 led to feelings of abandonment and subsequent despair. For example, many described that their loved one could not understand why visits were no longer being made and why they could not have physical contact. The following tweets illustrate these issues:

> *Or live alone with dementia and all the trouble I have. I can't even drive myself to a doctor. I don't remember all the rules myself. I'm terrible at wearing a mask. Someone pointed out I had it inside out at the covid [sic] testing place. Im gonna [sic] die, I hope. Im [sic] tired of life.*

> *You are told your Covid [sic] positive Mum is being discharged to you Covid [sic] negative dad. She has dementia. He is told to keep 2m away in their 3 bed bungalow. Is this NHS policy?*

> *I've lost count of how many times I've been on the phone with my grandmother telling her that it's not safe for her to leave the house because of COVID-19. Life with a loved one who has dementia right now is frustrating. Constantly re-teaching and remaining patient.*

> *Hardest thing to hear is my mom trying to explain to my grandmother, who has Alzheimer and dementia, that we can't see her because of the corona [sic]. My grandmother repeating that she is in jail. Asking where we are.*

Stress and Exhaustion Exacerbation

Care partners described stress and exhaustion (eg, mental, emotional, and physical) related to providing care for people with dementia in the context of the COVID-19 pandemic. Increased workload, disruption to routine, financial strain, mental health issues, social isolation, and loneliness were common features described.

COVID-19 confinement and lockdown measures substantially increased the workload of informal care partners by limiting or terminating access to support services such as day care programs, home care, respite, meal programs, medical specialists, and adequate care home options. As such, many described the difficulty of dealing with household chores, social isolation, and the increasing workload, which often led to feelings of mental, emotional, and physical exhaustion. Moreover, care partners described difficulties managing behavioral changes and worsening neuropsychiatric symptoms (eg, anxiety, agitation, anger, and depression) of their loved ones with dementia.

Care partners also discussed stress related to the fear of COVID-19 exposure and concern for the person with dementia's health and well-being. This fear was especially apparent among care partners with loved ones staying in care homes or hospitals. Others noted that COVID-19 precautions were confusing to people with dementia and made appointments and outings more difficult not only for the person with dementia but also for the care partner. Some reported balancing decisions on whether to pursue health services based on how distressing the experience would be for their loved one (eg, due to being alone without care partner support). Many reported strain due to financial pressures related to losing work coupled with unease for the future of their loved one. Difficulty obtaining care services or relocation to long-term care due to the pandemic were noted. Further, some care partners described feeling a need to bring people with dementia home from long-term care to ensure their well-being. The following tweets highlight the challenges faced by informal care partners:

I've lost some income but the hardest part is being stuck 24/7 in the house with my Mom who has severe dementia. It is hell! She had a daycare but it closed. I can't get respite bc [sic] I'm terrified of exposing her to covid [sic]. I fight loneliness, depression and boredom everyday [sic]!

The services for people with dementia have had to fall into line with COVID rules and there are of course personal restrictions. A devastating combination for people who need safety, routine and gentle stimulation.

Another horrifying day. We are in isolation with my beloved 93 yo [sic] Mom. She has descended into terrifying hallucinations and extreme anger because of dementia. We can't get her into nursing care because of a Covid [sic] outbreak. Just getting thru [sic] each hour.

Unpaid Sacrifices by Formal Care Providers

Formal care providers identified numerous sacrifices beyond their paid jobs. Formal care providers commonly expressed emotional connection to people with dementia and a sense of duty related to care, noting that this was more than simply a job. Formal care providers also described making personal sacrifices to work and provide care during the COVID-19 pandemic. For example, care providers discussed sacrificing their participation in family gatherings, parenting responsibilities, and social activities to help protect their patients and family members from potential exposure to the virus. Consequently, many expressed concern for the health of their families due to exposures at work and made trade-offs to help ensure the safety of their patients or residents, such as limiting their outside contacts. They also noted that caring for people with dementia involved a sacrifice of their own safety because of the lack of personal protective equipment (eg, gowns, gloves, masks, eye protection, and face shields), leading to stress of exposure. Finally, they reported their frustration with the reduced ability to support quality of life for people with dementia in the face of an increased workload due to COVID-19.

The workforce challenges experienced by formal care providers are captured in the following tweets:

People were discharged from hospital with covid [sic] and placed directly into my work. Trying to isolate a dementia patient with covid [sic] is impossible in a care home, they've already wandered down the corridor before we could even get our gloves on

I'm a mental health nurse working in a dementia specialist nursing home. My fight is to keep corona [sic] out of the building. There are many of us who will be in hiding to protect our residents…

I'm a nurse with COVID, probably from reusing dirty N95s and working with dementia patients who simply could not grasp the need to wear a mask and social distance. I worked so hard to try not to get COVID.

Discussion

Principal Results

Using Twitter, the aim of this study is to understand the COVID-19 experiences of people with dementia and their care partners. People with dementia are among the most vulnerable to the pandemic in terms of exposure risk, social isolation, hospitalization, and mortality [4]. Understanding the impact of COVID-19 is urgently needed to reduce mortality and improve the quality of life for people with dementia and their care partners during the pandemic. Given the current COVID-19 context, Twitter provided a valuable means to support timely and innovative research during the pandemic.

In analyzing the 5063 tweets, this study found that people with dementia are experiencing substantial burden from the COVID-19 pandemic. For example, separation and institutional visiting bans were described as having a detrimental impact on people with dementia. Numerous tweets identified the effects of visitation bans on people with dementia, such as despair, loss, abandonment, social isolation, not eating, losing the will to live, and dying alone. Tweets also identified challenges faced by informal care partners, such as financial struggles, mental health issues, lack of formal supports, inadequate care home options, fear of COVID-19 exposure, and difficulties explaining COVID-19 (eg, quarantine, self-isolation, social distancing, and protective equipment) to people with dementia. In addition, tweets addressed workforce issues experienced by formal care providers, ranging from insufficient access to personal protective equipment to understaffing and having to sacrifice family responsibilities to provide formal care.

This study has significant implications for COVID-19 policy and practice. For example, findings from this study suggest that prohibitive visitation policies (eg, visitation bans and institutional lockdowns) in care homes and health facilities may not be advantageous or acceptable for people with dementia. Numerous tweets emphasized the issue of separation on mortality for people with dementia. Emerging data from the United Kingdom showed that, in care homes, more people have died from dementia than from COVID-19 [2]. Similarly, data from Canada [17,18] and the United States [1] mirror this trend, documenting thousands of excess dementia deaths in care homes throughout the pandemic. These findings suggest that secondary effects of the COVID-19 pandemic (eg, social isolation) may be causing rapid deterioration and mortality of people with dementia. Consequently, COVID-19 policies (eg, visitation bans and institutional lockdowns) that were initially intended to protect people from the virus may be causing significant harm to people with dementia. Accordingly, there is an imminent need for governments to rethink using a one-size-fits-all response to COVID-19 policy and use instead a collaborative approach to support people with dementia and their care partners. In moving forward, collaboration and partnerships with people with dementia are essential to developing targeted policies to protect people with dementia and their care partners.

Formal care providers desperately need additional resources (eg, allocated funding, personal protective equipment, adequate staffing, and mental health supports) to support people with

dementia during the pandemic. Findings from this study identify a range of workforce challenges such as understaffing and lack of adequate personal protective equipment. Formal care providers in care homes and health facilities cannot provide optimal care and prevent COVID-19 exposure without having access to required resources and supports. Consequently, additional resources are needed to provide safeguards to protect both formal care providers and people with dementia during the pandemic.

Research is needed on the COVID-19 impact from the perspective of people with dementia and their care partners. Although several editorials, letters to the editor, and commentaries have discussed anticipated challenges of COVID-19 for people with dementia [8-10,19-21], few studies have involved people with dementia. Research is needed to examine the lived experiences of COVID-19 among people with dementia. Accordingly, more evidence-informed research is required to reduce mortality and understand the impact of COVID-19 on people with dementia and their care partners.

Limitations

Twitter has a 280-character limit for each tweet. Given this limit, the user's story and experience is confined to these restrictions. For example, important information and relevant details (eg, context, background, and confounding factors) may not be captured in the tweet. Consequently, qualitative interviews may provide a more comprehensive and in-depth perspective of the COVID-19 impact on people with dementia. In addition, given that Twitter is in the public domain, people may not feel comfortable sharing their full perspectives and lived experiences. Future research requires collaboration and partnerships with people with dementia to develop more in-depth knowledge on the impact of COVID-19 on people with dementia and their care partners.

In addition, there are some limitations related to the generalizability of our findings from Twitter. For example, no demographic information was collected in this study, and findings from Twitter may not be fully generalizable or representative of the general public living with dementia and their care partners (eg, age, gender, ethnicity, education, income, or employment background). However, existing data show that Twitter users are 56% male and 44% female, with the largest age groups of users being in the categories of 18-29 years (38%), 30-49 years (27%), and 50-64 years (17%) [22]. Since no demographic information was collected, another study limitation relates to sex and gender. Without any demographic information, it is impossible to make any inferences or draw specific conclusions regarding the impact of COVID-19 in relation to sex or gender. As such, more research is needed to examine the impact of COVID-19 in relation to the sex and gender of people with dementia and their care partners.

A final limitation of these data is the cross-sectional nature of our analysis and geographically blind nature of the data. It is possible that experiences of people with dementia and their care partners will vary over time, particularly over waves of the pandemic. We believe these current data captured what might be described as wave one of the pandemic. It is likely, however, that any temporal variability in experiences with COVID-19 and dementia will be related to geography. It is clear that there are geographic differences in the experiences of the pandemic. Future research to compare temporal variability in experiences of COVID-19 and dementia with publicly available databases detailing country-level COVID-19 disease burden and mitigation strategies is needed.

Conclusions

The purpose of this study was to use Twitter to understand the COVID-19 experiences of people with dementia and their care partners. Through an analysis of 5063 tweets, this study found that people are experiencing substantial burden from the COVID-19 pandemic. Specifically, four main issues were identified, including separation and loss, despair and abandonment, informal care partner challenges, and workforce issues experienced by formal care providers.

There is an imminent need for governments to rethink using a one-size-fits-all response to COVID-19 policy and use a collaborative approach to support people with dementia. More specifically, collaboration and partnerships are essential to developing effective COVID-19 policies to support people with dementia and their care partners. Moreover, there is a critical need for additional resources (eg, personal protective equipment, adequate staffing, safeguards, and mental health supports) to address workforce challenges and support formal care providers of people with dementia during the pandemic. Research is needed on the impact of COVID-19 from the perspective of people with dementia. Collaboration and more evidence-informed research are essential to reducing COVID-19–related mortality and improving the quality of lives for people with dementia and their care partners.

Acknowledgments

JDB is a postdoctoral fellow, KG is a doctoral student, and MA and SG were research assistants whose salaries were partially funded by Team 15 in the Canadian Consortium on Neurodegeneration in Aging (CCNA). CCNA is supported by a grant from the Canadian Institutes of Health Research with funding from several partners including the Saskatchewan Health Research Foundation, the Centre for Aging and Brain Health, and the Alzheimer Society of Canada (ASC). The ASC COVID Task Force (of which MEO, LP, and SS are members) was a particular catalyst for this study.

Authors' Contributions

MEO conceived the idea with LP. MEO, MA, and RJS devised the Twitter scraping approach, and MA scraped Twitter. MEO and JDB devised the codebook. MEO, AC, MA, KG, LP, SG, and RJS coded tweets, JDB double-coded a random sample of

these for quality control, and MEO was the final arbiter. All authors contributed example tweets. MEO and AC performed the thematic analysis. JDB wrote the first draft of the manuscript, and all authors revised the manuscript.

Conflicts of Interest

None declared.

References

1. Suárez-González A, Livingston G, Low LF, Cahill S, Hennelly N, Dawson WD, et al. Impact and mortality of COVID-19 on people living with dementia: cross-country report. International Long-Term Care Policy Network. 2020. URL: https://ltccovid.org/wp-content/uploads/2020/08/International-report-on-the-impact-of-COVID-19-on-people-living-with-dementia-19-August-2020.pdf [accessed 2020-10-19]
2. ONS report shows 52% increase in excess deaths of people dying of dementia – Alzheimer's Society comments. Alzheimer's Society. 2020. URL: https://www.alzheimers.org.uk/news/2020-06-05/ons-report-shows-52-increase-excess-deaths-people-dying-dementia-alzheimers-society [accessed 2020-10-23]
3. COVID-19 deaths disproportionately affecting people with dementia. Alzheimer's Disease International. 2020. URL: https://www.alzint.org/news/covid-19-deaths-disproportionately-affecting-people-with-dementia/ [accessed 2020-10-23]
4. Azarpazhooh MR, Amiri A, Morovatdar N, Steinwender S, Rezaei Ardani A, Yassi N, et al. Correlations between COVID-19 and burden of dementia: an ecological study and review of literature. J Neurol Sci 2020 Sep 15;416:117013 [FREE Full text] [doi: 10.1016/j.jns.2020.117013] [Medline: 32659508]
5. Older adults. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html [accessed 2020-10-21]
6. National Academies of Sciences, Engineering, and Medicine, Division of Behavioral and Social Sciences and Education, Health and Medicine Division, Board on Behavioral, Cognitive, and Sensory Sciences, Board on Health Sciences Policy, Committee on the Health and Medical Dimensions of Social Isolation and Loneliness in Older Adults. Social Isolation and Loneliness in Older Adults: Opportunities for the Health Care System. Washington, DC: National Academies Press; 2020.
7. Vaitheswaran S, Lakshminarayanan M, Ramanujam V, Sargunan S, Venkatesan S. Experiences and needs of caregivers of persons with dementia in India during the COVID-19 pandemic-a qualitative study. Am J Geriatr Psychiatry 2020 Nov;28(11):1185-1194 [FREE Full text] [doi: 10.1016/j.jagp.2020.06.026] [Medline: 32736918]
8. Wang H, Li T, Barbarino P, Gauthier S, Brodaty H, Molinuevo JL, et al. Dementia care during COVID-19. Lancet 2020 Apr 11;395(10231):1190-1191 [FREE Full text] [doi: 10.1016/S0140-6736(20)30755-8] [Medline: 32240625]
9. Canevelli M, Bruno G, Cesari M. Providing simultaneous COVID-19-sensitive and dementia-sensitive care as we transition from crisis care to ongoing care. J Am Med Dir Assoc 2020 Jul;21(7):968-969 [FREE Full text] [doi: 10.1016/j.jamda.2020.05.025] [Medline: 32536553]
10. Migliaccio R, Bouzigues A. Dementia and COVID-19 lockdown: more than a double blow for patients and caregivers. J Alzheimers Dis Rep 2020 Jul 06;4(1):231-235 [FREE Full text] [doi: 10.3233/ADR-200193] [Medline: 32904665]
11. Clay RA. Conducting research during the COVID-19 pandemic. American Psychological Association. 2020. URL: https://www.apa.org/news/apa/2020/03/conducting-research-covid-19 [accessed 2020-10-23]
12. Clement J. Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019. Statista. 2019. URL: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ [accessed 2020-10-19]
13. Twitter. URL: https://twitter.com/ [accessed 2020-10-19]
14. Henrique J. A project written in Python to get old tweets which could bypass some limitations of Twitter official API. GitHub. URL: https://github.com/Jefferson [accessed 2020-10-21]
15. Moreno MA, Goniu N, Moreno PS, Diekema D. Ethics of social media research: common concerns and practical considerations. Cyberpsychol Behav Soc Netw 2013 Sep;16(9):708-713 [FREE Full text] [doi: 10.1089/cyber.2012.0334] [Medline: 23679571]
16. Talbot CV, O'Dwyer ST, Clare L, Heaton J, Anderson J. How people with dementia use twitter: a qualitative analysis. Comput Hum Behav 2020 Jan;102:112-119. [doi: 10.1016/j.chb.2019.08.005]
17. Hsu AT, Lane N, Sinha SK, Dunning J, Dhuper M, Kahiel Z, et al. Impact of COVID-19 on residents of Canada's long-term care homes – ongoing challenges and policy responses. International Long-Term Care Policy Network. 2020. URL: https://ltccovid.org/wp-content/uploads/2020/05/LTCcovid-country-reports_Canada_Hsu-et-al_May-10-2020-2.pdf [accessed 2020-10-23]
18. McMichael TM, Currie DW, Clark S, Pogosjans S, Kay M, Schwartz NG, Public Health–Seattle and King County, EvergreenHealth, and CDC COVID-19 Investigation Team. Epidemiology of Covid-19 in a long-term care facility in King County, Washington. N Engl J Med 2020 May 21;382(21):2005-2011 [FREE Full text] [doi: 10.1056/NEJMoa2005412] [Medline: 32220208]
19. Brown EE, Kumar S, Rajji TK, Pollock BG, Mulsant BH. Anticipating and mitigating the impact of the COVID-19 pandemic on Alzheimer's disease and related dementias. Am J Geriatr Psychiatry 2020 Jul;28(7):712-721 [FREE Full text] [doi: 10.1016/j.jagp.2020.04.010] [Medline: 32331845]

20. Korczyn AD. Dementia in the COVID-19 period. J Alzheimers Dis 2020;75(4):1071-1072 [FREE Full text] [doi: 10.3233/JAD-200609] [Medline: 32538858]
21. Velayudhan L, Aarsland D, Ballard C. Mental health of people living with dementia in care homes during COVID-19 pandemic. Int Psychogeriatr 2020 Oct;32(10):1253-1254 [FREE Full text] [doi: 10.1017/S1041610220001088] [Medline: 32487278]
22. Clement J. Percentage of U.S. adults who use Twitter as of February 2019, by age group. Statista. URL: https://www.statista.com/statistics/265647/share-of-us-internet-users-who-use-twitter-by-age-group/ [accessed 2020-10-19]

Abbreviations

ASC: Alzheimer Society of Canada
CCNA: Canadian Consortium on Neurodegeneration in Aging

Edited by G Eysenbach; submitted 03.12.20; peer-reviewed by B Chakalov, C Basch; comments to author 24.12.20; accepted 19.01.21; published 03.02.21.

Please cite as:
Bacsu JD, O'Connell ME, Cammer A, Azizi M, Grewal K, Poole L, Green S, Sivananthan S, Spiteri RJ
Using Twitter to Understand the COVID-19 Experiences of People With Dementia: Infodemiology Study
J Med Internet Res 2021;23(2):e26254
URL: https://www.jmir.org/2021/2/e26254
doi:10.2196/26254
PMID:33468449

Original Paper

Factors Driving Citizen Engagement With Government TikTok Accounts During the COVID-19 Pandemic: Model Development and Analysis

Qiang Chen[1], PhD; Chen Min[2,3], MA; Wei Zhang[4], PhD; Xiaoyue Ma[1], PhD; Richard Evans[5], PhD

[1]School of Journalism and New Media, Xi'an Jiaotong University, Xi'an, China

[2]Department of Media and Communication, City University of Hong Kong, Hong Kong, China

[3]College of Public Administration, Huazhong University of Science and Technology, Wuhan, China

[4]School of Medicine and Health Management, Huazhong University of Science and Technology, Wuhan, China

[5]College of Engineering, Design and Physical Sciences, Brunel University London, London, United Kingdom

Corresponding Author:
Wei Zhang, PhD
School of Medicine and Health Management
Huazhong University of Science and Technology
13 Hangkong Road
Wuhan, 430074
China
Phone: 86 13397110378
Email: weizhanghust@hust.edu.cn

Abstract

Background: During the COVID-19 pandemic, growth in citizen engagement with social media platforms has enabled public health departments to accelerate and improve health information dissemination, developing transparency and trust between governments and citizens. In light of these benefits, it is imperative to learn the antecedents and underlying mechanisms for this to maintain and enhance engagement.

Objective: The aim of this study is to determine the factors and influencing mechanisms related to citizen engagement with the TikTok account of the National Health Commission of China during the COVID-19 pandemic.

Methods: Using a web crawler, 355 short videos were collected from the Healthy China account on TikTok (with more than 3 million followers throughout China), covering the period from January 21, 2020, to April 25, 2020. The title and video length, as well as the number of likes, shares, and comments were collected for each video. After classifying them using content analysis, a series of negative binomial regression analyses were completed.

Results: Among the 355 videos, 154 (43.4%) related to guidance for clinicians, patients, and ordinary citizens, followed by information concerning the government's handling of the pandemic (n=100, 28.2%), the latest news about COVID-19 (n=61, 17.2%), and appreciation toward frontline emergency services (n=40, 11.3%). Video length, titles, dialogic loop, and content type all influenced the level of citizen engagement. Specifically, video length was negatively associated with the number of likes (incidence rate ratio [IRR]=0.19, *P*<.001) and comments (IRR=0.39, *P*<.001). Title length was positively related to the number of shares (IRR=24.25, *P*=.01), likes (IRR=8.50, *P*=.03), and comments (IRR=7.85, *P*=.02). Dialogic loop negatively predicted the number of shares (IRR=0.56, *P*=.03). In comparison to appreciative information, information about the government's handling of the situation (IRR=5.16, *P*<.001) and guidelines information (IRR=7.31, *P*<.001) were positively correlated with the number of shares, while the latest news was negatively related to the number of likes received (IRR=0.46, *P*=.004). More importantly, the relationship between predictors and citizen engagement was moderated by the emotional valence of video titles. Longer videos with positive titles received a higher number of likes (IRR=21.72, *P*=.04) and comments (IRR=10.14, *P*=.047). Furthermore, for short videos related to government handling of the pandemic (IRR=14.48, P=.04) and guidance for stakeholders (IRR=7.59, *P*=.04), positive titles received a greater number of shares. Videos related to the latest news (IRR=66.69, *P*=.04) received more likes if the video title displayed higher levels of positive emotion.

Conclusions: During the COVID-19 pandemic, videos were frequently published on government social media platforms. Video length, title, dialogic loop, and content type significantly influenced the level of citizen engagement. These relationships were

moderated by the emotional valence of the video's title. Our findings have implications for maintaining and enhancing citizen engagement via government social media.

(J Med Internet Res 2021;23(2):e21463) doi:10.2196/21463

KEYWORDS

government social media; citizen engagement; public health crisis; TikTok; emotion valence; dialogic loop; COVID-19

Introduction

Background

Citizen engagement on social media during times of public health crises, such as that experienced during the COVID-19 pandemic, provides governments with a valuable tool for communicating with and understanding the concerns and priorities of their citizens [1]. Citizen engagement highlights the active role citizens play in public affairs, including public communication, public consultation, and public participation, which have the potential to influence government decision making [2]. With citizen engagement, governments can solicit public concerns and respond accordingly, improving the quality and efficiency of public services delivered [3]. Due to the popularity and communicative nature of social media, it has been employed widely by governments during public health crises, creating new opportunities for citizens to voluntarily participate in government activities [4-8]. However, the practice of engaging citizens is relatively underdeveloped. Strategies for citizen engagement via social media platforms have not been adequately deployed, although the importance of citizen engagement is widely acknowledged by government agencies. The use of social media by governments during public crises is hampered by both internal and external factors, such as insufficient resources, the digital divide, ethical considerations, and accountability [9,10]. Government agencies thus prefer to broadcast information and improve disease surveillance through social media, rather than initiate public conversation and engagement [11-13]. Considering its benefits, it is imperative to better understand the antecedents and underlying mechanisms of citizen engagement with government social media during public health emergencies to maintain and enhance citizen engagement.

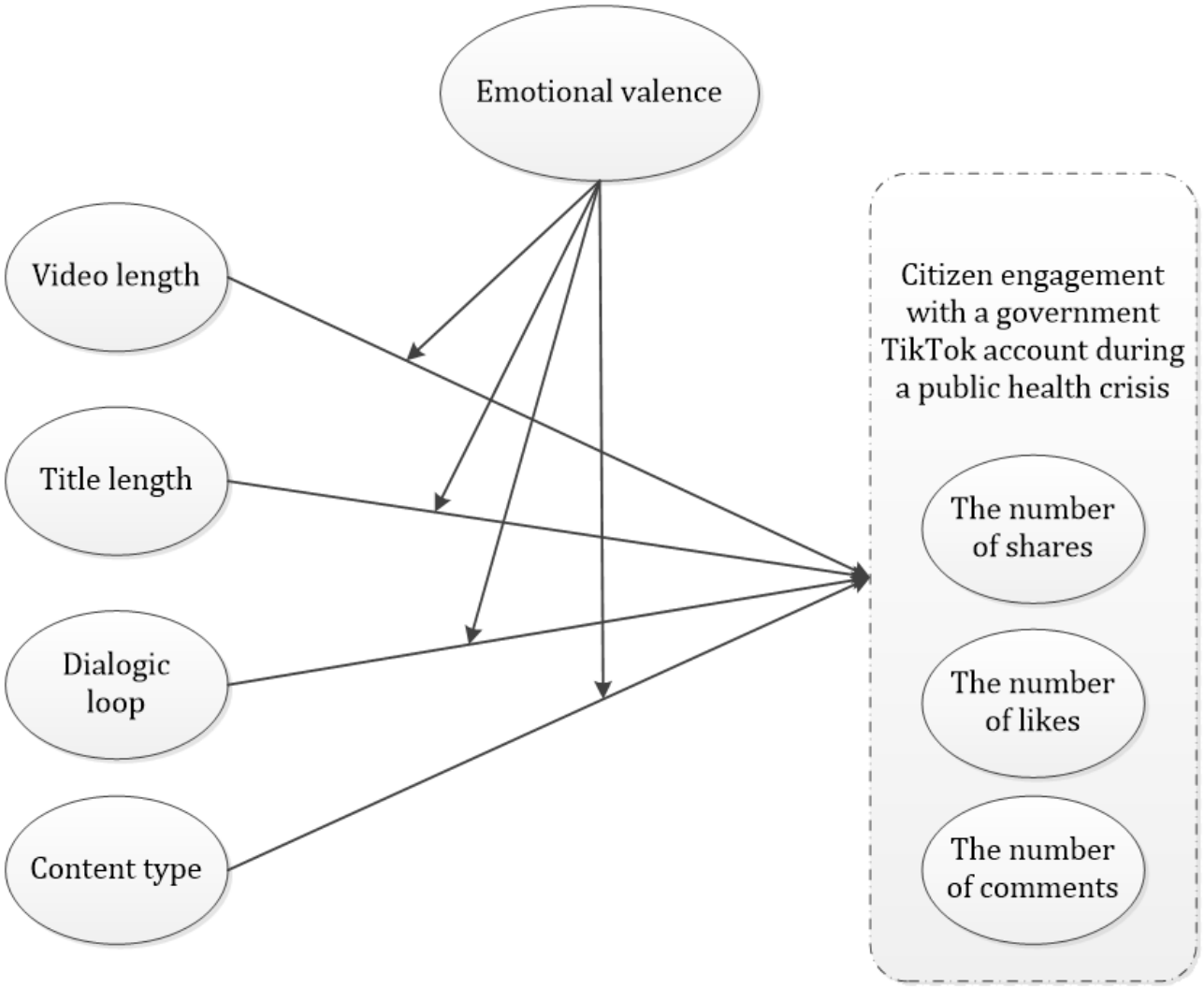
Although some studies have examined the relationship between government agencies' use of social media and citizen engagement during public health crises, there are several obvious gaps that deserve further investigation. First, existing studies have paid extensive attention to more traditional social media, including Sina Weibo [1,8], Twitter [4,14], and Facebook [7]. Chen and colleagues [14] identified the tweeting patterns of the Centers for Disease Control and Prevention during different phases of the Zika epidemic. However, their studies focused on text-dominated social media exclusively, with little attention paid to emerging social media platforms. For example, TikTok, which launched in 2016, is a video-sharing platform that allows users to create, upload, repost, and write comments on short videos, ranging from a few seconds to minutes. Since its creation, the popularity of TikTok has increased rapidly, with more than 800 million active users in 2020 [15]. During the COVID-19 pandemic, TikTok became an indispensable source of information and communication channel for the general public and between government organizations and citizens [16]. Second, extant studies have predominantly explored possible influencing factors through the lens of media type, government type, content type, media richness, message style, and interactive features, with limited effort paid to specific influencing mechanisms [1,4,8]. Guidry et al [4] explored the Ebola-related social media posts published by three major health organizations on Twitter and Instagram, and found that messages with solutions, visual imagery, and acknowledgment of public concerns are the most effective risk communication strategies. By investigating the People's Daily Sina Weibo account during COVID-19, Ngai et al [8] found that disease prevention content posts in a narrative style generated more public engagement, and further revealed an interaction effect between content and style.

This study aims to address the above gaps by exploring the factors that drove citizen engagement through the TikTok account of the National Health Commission of China (NHCC) during the COVID-19 pandemic, in the context of China. Chinese governments, across different levels, have now started to realize the potential of TikTok for public participation, leveraging the platform for routine administration tasks. According to the latest report by the China Internet Network Information Center (CNNIC), the number of verified government TikTok accounts is 25,313, implying wide employment in the Chinese public sector [17]. In this study, we develop a conceptual model, including video length, title length, dialogic loop, and content type, to empirically examine the driving factors of citizen engagement with TikTok accounts operated by Chinese government agencies. Citizen engagement through the official TikTok account of the NHCC consists of three dimensions: comments, likes, and shares [18]. For each video uploaded to TikTok, data on its number of comments, likes, and shares are publicly available. Importantly, we take the emotional valence variable as a moderator to unravel potential influencing mechanisms.

Research Hypotheses and Conceptual Model

Based on the characteristics of TikTok videos, our study built a moderated model that unraveled the moderating role of emotional valence among the differentiated effects of video length, title length, dialogic loop, and content type on the three dimensions of citizen engagement, namely the number of shares, likes, and comments (Figure 1).

Figure 1. The research model of the study.



Video length refers to the duration of a video. Although several studies have examined the length of popular videos, few have investigated the impact of video length on citizen engagement on TikTok. For example, Zhu et al [19] analyzed the 100 most-liked videos on TikTok published by 31 provincial Health Committees in China, and found that 98% of them lasted under 60 seconds. Our study suggests that video length is negatively associated with the number of likes, shares, and comments received from the public. Shorter videos usually have better content because they require detailed planning to explain the content concisely [20]. Guo et al [20] analyzed a big data set comprising 6.9 million Massive Open Online Course (MOOC) videos and confirmed that shorter videos had a perceived higher quality, thus leading to increased engagement. In addition, shorter videos align better with the information consumption habits of social media users. Users usually watch TikTok videos on mobile devices during fragmented periods. If a video is too long, it will not only consume more mobile data and battery power, but also fail to meet the demands of individuals' allotted time.

Therefore, hypothesis 1 is as follows: video length negatively affects the three dimensions of citizen engagement via the official TikTok account of the NHCC, as measured by the number of likes, shares, and comments.

Title length refers to the total number of words summarizing the video's content in the title of an uploaded video. During times of crisis, citizens are usually motived by utility, craving information that relieves anxiety and eases panic [21]. Longer titles typically contain richer information, promoting a better understanding of the video, thereby increasing citizen engagement [21,22]. For example, Xu and Zhang [22] examined 13,322 tweets related to Malaysia Airlines Flight 370 and found that the more words a tweet had, the more likely it was to be retweeted. Specific to video-sharing platforms, Halvey and Keane [23] analyzed 4.25 million YouTube videos and identified that a longer video title increased the number of video views. We argue that the longer the title length, the more likely citizens are to share, comment, and like COVID-19–related TikTok videos.

Therefore, hypothesis 2 is as follows: title length positively affects the three dimensions of citizen engagement via the official TikTok account of the NHCC, as measured by the number of likes, shares, and comments.

Dialogic loop emphasizes the importance of providing an interactive feedback loop, facilitating mutual communication between organizations and citizens [1]. Organizations allow the general public to directly ask questions, post comments, make

suggestions, and provide feedback on relevant issues pertaining to the organization [24]. Dialogic loop is beneficial for promoting mutual conversation through question posing and answering [1]. However, few studies have investigated the effect of dialogic loop on citizen engagement via social media. Gálvez-Rodríguez et al [25] analyzed 137 Latin American local governments' Facebook accounts and revealed that dialogic loop can increase the level of citizen engagement (weighed calculation based on the number of likes, comments, shares, and posts). Specific to public health crises, Chen et al [1] studied the official Sina Weibo account of the NHCC and verified that dialogic loop positively influences citizen engagement (sum of number of likes, comments, and shares).

Therefore, hypothesis 3 is as follows: dialogic loop positively affects the three dimensions of citizen engagement via the official TikTok account of the NHCC, as measured by the number of likes, shares, and comments.

Extant studies into health communication have identified that different content types have varying effects on citizens' social media engagement behaviors. According to the Use and Gratification Theory, the degree to which media content gratifies individuals' differential needs directly influences their media selection and usage behavior [26]. Analyzing 203,191 COVID-19–related posts on Sina Weibo, Wang et al [27] concluded that posts on the domestic epidemic were retweeted the most, while posts about quarantine and investigations received the most likes. Short videos related to COVID-19, posted by the official TikTok account of the NHCC during the pandemic, will satisfy public demands to different degrees, thus promoting citizen engagement at different levels. Research that analyzed 4221 tweets posted by the health departments of 39 states in the United States showed that acknowledging the events of other organizations and expressing gratitude and providing recognition improved citizen engagement [28]. Park et al [29] examined 1583 tweets posted by three American health organizations and found that personal health–related information and actions received more shares and likes. In relation to public health crises, Chen et al [1] determined that posts about the latest news and government disposition facilitated citizen engagement via government Sina Weibo accounts, while posts providing guidance for stakeholders had no effect.

Therefore, hypothesis 4 is as follows: content type has significant differential effects on the three dimensions of citizen engagement via the official TikTok account of the NHCC, as measured by the number of likes, shares, and comments.

Emotional valence refers to the positive and negative feelings triggered during an individual's consumption of information [1,18]. Emotional expression in social media content can attract public attention, promote dialogue, and drive engaged feedback [18]. Emotions possess the function of physiological arousal, which can contribute to information-sharing behavior [30]. High emotional traits can trigger the viral spread of social media content [22]. Citizens regulate their emotional status through actions such as reposting, liking, and commenting when receiving emotional information via social media [1]. Empirical evidence indicates that emotional posts often promote citizens' social media engagement [18,31,32]. However, few studies have investigated the moderating effect of emotional valence [33]. Tang et al [34] studied the official Sina Weibo accounts of 30 provincial police departments in China and identified that emotional valence moderated the influence of content type on reposting behavior. Child-friendly content with positive emotion increased reposting the most. Current research has also confirmed that emotional valence moderates the effects of dialogic loop, media richness, and content type on citizen engagement via government Sina Weibo accounts during public health crises [1].

Therefore, hypothesis 5 is as follows: emotional valence will moderate the effects of video length, title length, dialogic loop, and content type on the three dimensions of citizen engagement via the official TikTok account of the NHCC, as measured by the number of likes, shares, and comments.

Methods

Data Collection

For this study, we collected data from the official TikTok account of the NHCC, "Healthy China." The official account was created on May 4, 2018, to help more widely disseminate health-related information. By May 12, 2020, the account had posted 576 short videos, had 3 million followers, and had received more than 8 million likes. Since the outbreak of COVID-19 in December 2019, the NHCC has actively uploaded relevant short videos. We used a web crawler tool to capture all videos uploaded to the official account from January 21, 2020, to April 25, 2020. In total, 364 videos were obtained. After manual checking, 355 were found to be related to the COVID-19 crisis. In addition, data about the video's length, title text, number of likes, number of shares, and number of comments were collected.

Operationalization of Variables

Citizen engagement through the official TikTok account of the NHCC includes three dimensions: sharing, liking, and commenting behaviors [18]. We captured this objective data using web crawlers.

Video length is the duration of the video. The length of all 355 videos was captured from the "Healthy China" account using a web crawler.

Title length relates to the number of words included in a video's title. This study first collected the complete texts of 355 video titles using web crawlers, and then automatically counted the number of words in each text using Microsoft Excel (Microsoft Corp).

Dialogic loop is measured by two indicators: posing and answering questions [35]. If a question or answer appears in a short video title, it is marked as 1, otherwise it is marked as 0. The dialogic loop score was the sum of the scores of the two indicators.

Content type includes 4 categories, namely the following: the latest news about the COVID-19 crisis, the government's handling of the pandemic, appreciation for frontline emergency services, and guidance for stakeholders [1]; the codebook in Table 1 contains examples.

Table 1. Codebook examples of content categories.

Categories	Example posts
The latest news about the COVID-19 crisis	[Authoritative Release] The latest situation of COVID-19 crisis as of 24:00 on April 11#Action to defeat COVID-19
The government's handling of the crisis	Covid-19 prevention, control and medical treatment work in Hubei, from the press conference of the State Council Information Office. #Action to defeat COVID-19
Appreciation of frontline emergency services	Thank you, Sichuan Provincial Medical Aid Team, tribute heroes triumphantly! #Action to defeat COVID-19 #Warm doctor moment
Guidance for stakeholders	Parents look out! How can children prevent COVID-19? #Action to defeat COVID-19

Emotional Valence

We measured this variable by calculating the emotional valence of each short video's title based on a sentiment lexicon and Python (Python Software Foundation). Initially, title information was split into three categories (sentimental words, negative words, and degree adverbs) using the Jieba database. Upon word separation, the words were annotated using BosonNLP, while machine learning enabled word processing to assign values for sentimental words. By combining the review of the rest of the words with preassigned values, the emotional value for each title was determined based on a weighted approach. Each title was assigned a value ranging from 0 to 1, with 0.5 being neutral emotion. The closer the value was to 0, the more negative the emotion was considered, and vice versa.

Intercoder Reliability and Data Analysis

This study obtained the analysis data of dialogic loop and content type by manually coding the content of the video titles. We employed two graduate students to carry out the coding of the 355 short videos. Before beginning, the students were provided with operation training and an explanation of the coding scheme to ensure a good coding standard. To examine the interreliability, 30% of the sample was randomly selected, and 106 short videos were precoded. Both coders worked independently, and the results are as follows: the κ values for the content categories, "posting a question," and "responding to a question" were 0.918, 0.941, and 0.795, respectively. This indicated that the interreliabilities were high enough and acceptable.

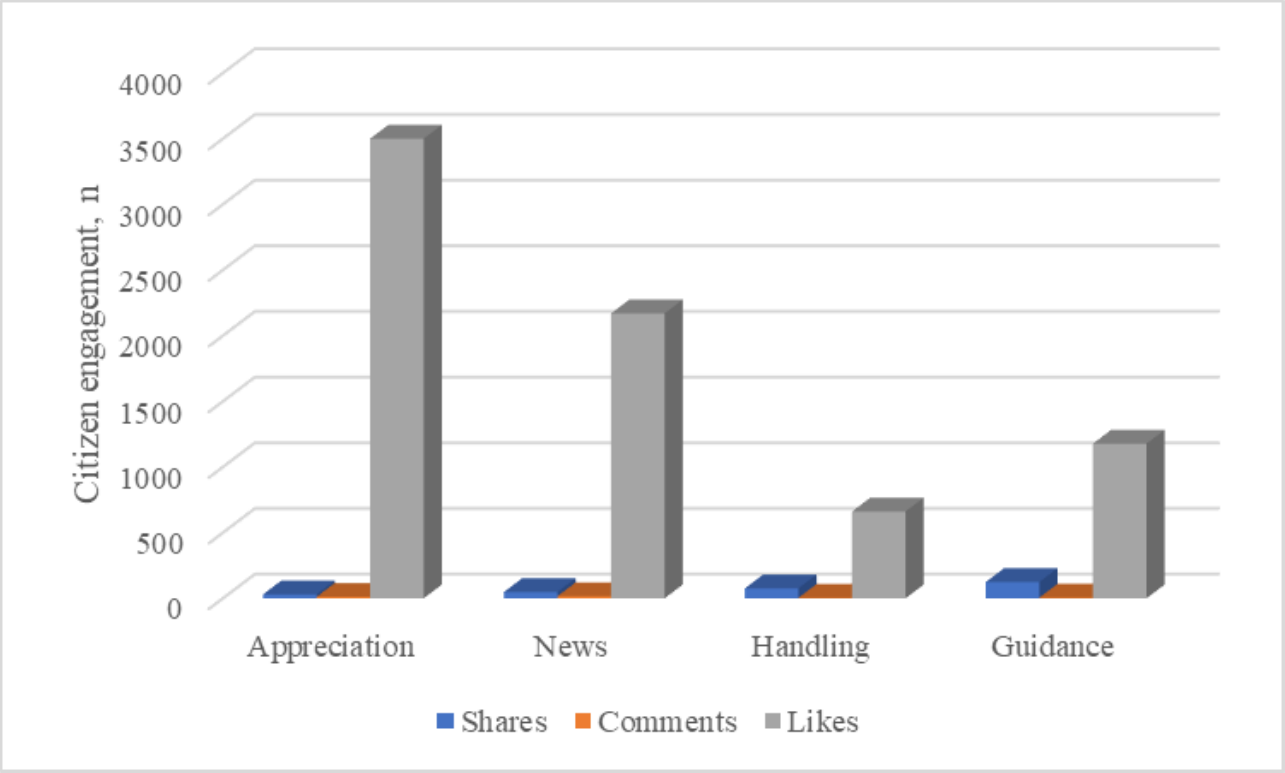
Our dependent variables are all count measures that exhibit overdispersion (shares: mean 86.94 [SD 492.71], skewness=11.36, kurtosis=140.27; likes: mean 1464.22 [SD 4772.29], skewness=6.94, kurtosis=57.85; comments: mean 7.57 [SD 13.22], skewness=4.10, kurtosis=26.07). A significant number of short videos rarely received shares, likes, or comments, while others received a high number in comparison. To deal with this overdispersed count data, we modelled the number of shares, likes, and comments using negative binomial regression. We first estimated the impact of video length, title length, dialogic loop, and content type on different types of citizen engagement. We then explored whether the impacts were contingent upon the emotional valence of each video's title. All analyses were conducted using STATA (Version 15.0; StataCorp LLC).

Results

Descriptive Analysis

The 355 collected short videos had large variations in the level of citizen engagement. Overall, 32.1% of videos (n=114) had less than 10 shares, while 9.9% of videos (n=35) were shared more than 100 times. The gap in the number of likes was also significant, with 56 videos (15.8%) receiving less than 100 likes and 10 videos (2.8%) receiving more than 10,000 likes. For comments, 8 videos received more than 50 comments, while 18.9% of videos (n=67) did not receive any comments. Among all videos, 154 (43.4%) were related to guidance for clinicians, patients, and ordinary citizens, followed by information concerning the government's handling of the pandemic (n=100, 28.2%), and the latest news about COVID-19 (n=61, 17.2%). Posts about appreciation toward frontline emergency services only represented about 11.3% (n=40) of posts. On average, the appreciation posts received the highest number of likes (mean 3500.28 [SD 8086.11]), while guidance and government handling information received the least (mean 660.97 [SD 1493.57] and mean 1177.60 [SD 3952.23], respectively; Figure 2). Moreover, videos related to guidance information experienced the highest level of sharing (mean 124.12 [SD 624.95]), while videos pertaining to the latest news had the highest number of comments (mean 16.07 [SD 21.44]). Of the other characteristics of videos, the length of all videos collected ranged from 11 seconds to 493 seconds (mean 96.53 [SD 81.78]). Furthermore, the average length of video titles was 34.75 characters (SD 6.79; range 15-54). For dialogic loop, 203 videos (57.2%) either raised or answered questions to deliver information. Further, the emotional valence of video titles was positive on average (mean 0.71 [SD 0.30]).

Figure 2. Average amount of citizen engagement in the form of shares, likes, and comments, grouped by content type.



Hypothesis Testing

Table 2 shows estimates of the negative binomial regression models that predicted the number of likes, shares, and comments for each coded short video uploaded during the COVID-19 pandemic. Hypothesis 1 posited that longer videos were less likely to attract citizen engagement. Models 2 and 3 in Table 2 show that video length has a negative and statistically significant association with the number of likes (incidence rate ratio [IRR]=0.19, *P*<.001) and comments (IRR=0.39, *P*<.001) received. The IRR value shows that a one-unit increase in video length leads to a decrease in the number of likes and comments by a factor of 0.19 and 0.39, respectively. However, the relationship between video length and number of shares is not significant. Thus, hypothesis 1 was partially supported.

Table 2. Predicting citizen engagement with government social media.

Variables	Shares (Model 1)		Likes (Model 2)		Comments (Model 3)	
	Incidence rate ratio	*P* value	Incidence rate ratio	*P* value	Incidence rate ratio	*P* value
Video length	0.81	.59	0.19	<.001	0.39	<.001
Title length	24.25	.01	8.50	.03	7.85	.02
Dialogic loop	0.56	.03	0.67	.07	0.68	.07
Content type (reference: appreciation)						
News	1.15	.67	0.46	.004	0.99	.96
Handling	5.16	<.001	0.97	.94	1.06	.86
Guidelines	7.31	<.001	0.76	.36	.89	.67
Constant	0.28	.51	1398.47	<.001	1.85	.68
Log likelihood	–1797.08	N/A[a]	–2816.56	N/A	–1026.53	N/A
Pseudo R^2 (%)	1.41	N/A	1.67	N/A	3.68	N/A

[a]N/A: not applicable.

Hypothesis 2 proposed that a longer video title increases the level of citizen engagement. According to Table 2, title length indeed plays an important role in predicting the number of shares (IRR=24.25, *P*=.01), likes (IRR=8.50, *P*=.03), and comments (IRR=7.85, *P*=.02). A one-unit increase in title length can result in a 2325% increase in the number of shares, a 750% increase in the number of likes, and a 685% increase in the number of comments. Thus, hypothesis 2 was supported.

Hypothesis 3 proposed that videos with multiple dialogic features were more likely to attract citizen engagement. Model 1 in Table 2 shows that dialogic loop is negatively associated with the number of shares (IRR=0.56, *P*=.03). This means that dialogic loop decreases willingness to share by about 44%. In addition, Models 2 and 3 show that dialogic loop is not associated with the number of likes and comments received. Thus, hypothesis 3 was not supported.

Hypothesis 4 proposed that the degree of citizen engagement was influenced by the content shown in the videos. Since content type is a categorical variable, we treated appreciative posts as the reference group. Model 1 in Table 2 shows that government handling information (IRR=5.16, *P*<.001) and guidelines information (IRR=7.31, *P*<.001) were positively correlated with the number of shares. The IRR values mean that, compared to appreciative information, handling information would result in a 416% increase in the number of shares, while guideline information would lead to a 631% increase. Videos about the latest news were found to be negatively related to the number of likes (IRR=0.46, *P*=.004), which means that posts about the latest news would result in a 54% decrease in the number of likes received, compared to posts about appreciative information. Interestingly, the number of comments did not vary significantly across different types of video content. Thus, hypothesis 4 was partially supported.

We investigated the conditional impacts of the predictors by entering emotional valence and interaction variables into the negative binomial regression model. As seen in Table 3, the relationship between video length and citizen engagement is moderated by the emotional valence of a video's title. The interaction between video length and emotion is positively related to the number of likes (IRR=21.72, *P*=.04) and comments (IRR=10.14, *P*=.047; see Models 2 and 3). As Figure 3 shows, the gap in the number of likes between videos with positive emotion and negative emotion is only significant when the video length is long. The more negative the emotion in the video's title, the lower the number of likes received. This pattern is almost the same for the number of comments (Figure 4). The lowest number of comments occurred when the video was long and had negative emotion. Emotion also moderates the relationship between content type and the number of shares. As shown in Table 3 (Model 2), for guidelines (IRR=7.59, *P*=.04) and government handling (IRR=14.48, *P*=.04) information, the more positive emotion a video's title had, the higher the number of shares the video received, in comparison to appreciative news. Videos related to the latest news (IRR=66.69, *P*=.04) received more likes when their titles had higher levels of positive emotion. Interestingly, the title's emotional valence has a weak effect or no effect on the relationship between content and other types of engagement behavior. Overall, emotion plays a moderate role in the relationship between video length, content type, and citizen engagement, but patterns vary across different types of engagement.

Table 3. The moderating role of emotional valence in predicting citizen engagement.

Variables	Shares (Model 1)		Likes (Model 2)		Comments (Model 3)	
	Incidence rate ratio	*P* value	Incidence rate ratio	*P* value	Incidence rate ratio	*P* value
Video length	0.67	.28	0.12	<.001	0.28	<.001
Title length	32.48	.008	7.84	.08	3.69	.23
Dialogic loop	0.68	.15	0.70	.13	0.63	.03
Content type (reference: appreciation)						
News	2.36	.25	0.17	.02	2.03	.23
Handling	0.51	.49	2.03	.55	0.99	.99
Guidelines	1.15	.86	1.63	.65	3.02	.17
Interaction						
Emotion	1.10	.91	5.70	.05	4.50	.03
Emotion × video	0.45	.56	21.72	.04	10.14	.047
Emotion × title	1.60	.91	4.41	.77	0.60	.90
Emotion × content type (reference: appreciation)						
News	0.11	.19	66.69	.04	0.09	.06
Handling	14.48	.04	0.38	.50	0.98	.99
Guidelines	7.59	.04	0.37	.43	0.21	.10
Constant	0.25	.51	3313.61	<.001	9.84	.22
Log likelihood	–1785.57	N/A[a]	–2806.04	N/A	–1017.43	N/A
Pseudo R^2 (%)	2.04	N/A	2.03	N/A	4.53	N/A

[a]N/A: not applicable.

Figure 3. Interaction effects of video length and emotion on the number of likes received.

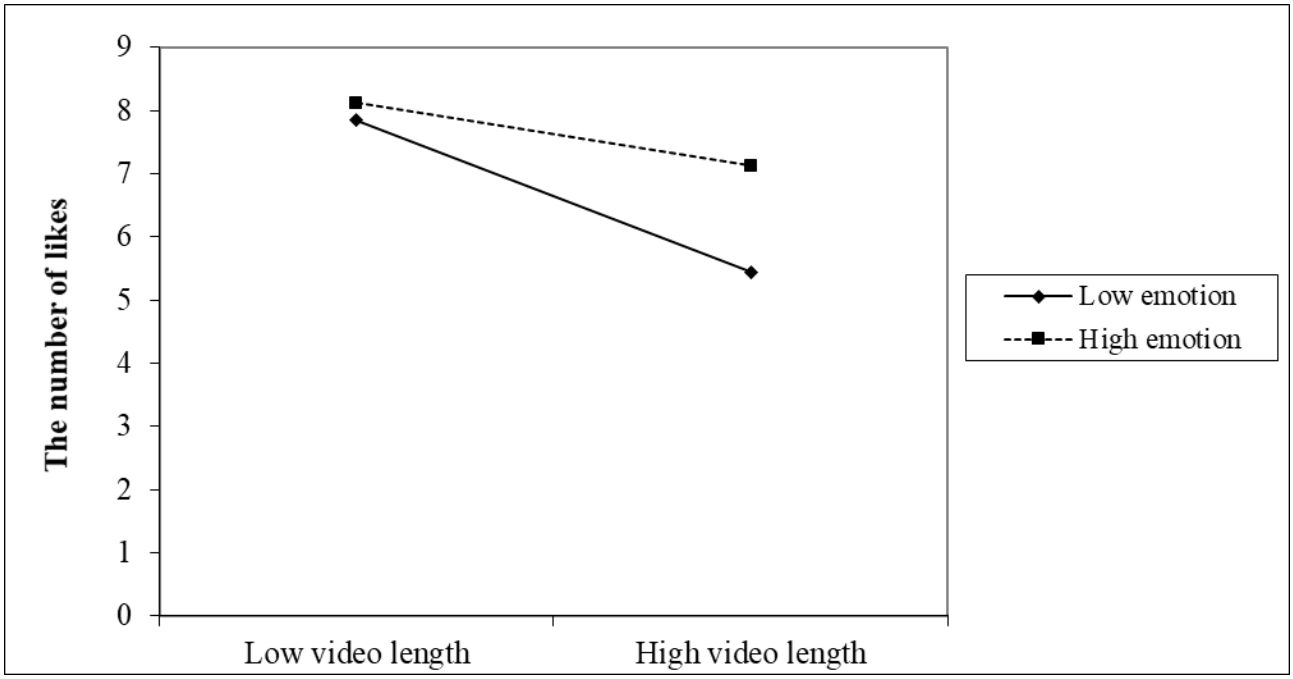
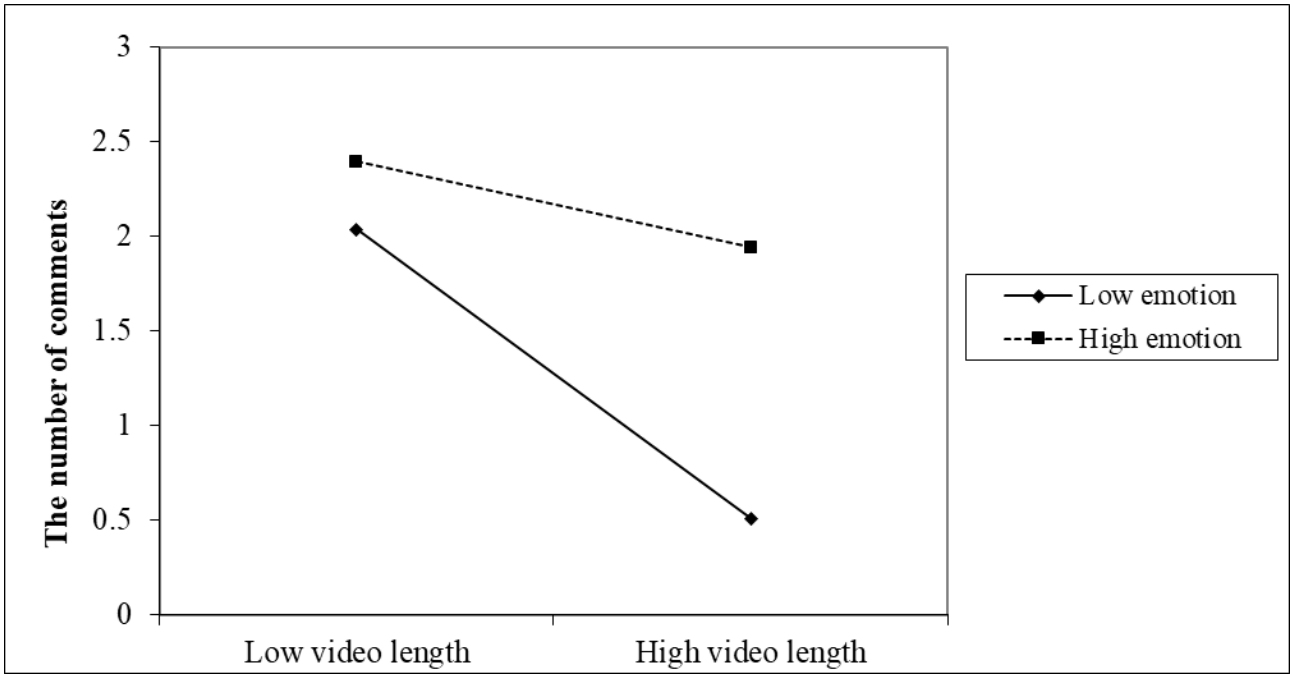


Figure 4. Interaction effects of video length and emotion on the number of comments received.



Discussion

Principal Findings

This pioneering study is the first to attempt to reveal how to promote citizen engagement with official government TikTok accounts during the COVID-19 crisis. To reveal the specific mechanisms of engagement, we proposed a moderated model, which investigated the moderating role of emotional valence on the effects of video length, title length, dialogic loop, and content type on the three dimensions of citizen engagement with the official TikTok account of the NHCC.

First, we found that shorter videos and videos with longer titles generate greater engagement in general. On the one hand, video length negatively and significantly influenced the number of likes and comments received. This conclusion confirms the findings of Zhu et al [19], who found that citizens prefer shorter TikTok videos. We extend their findings in the context of public health crises. Shorter videos can better satisfy public demands for timely consumption and relaxation during fragmented time and promote the public's continuous use of TikTok videos, including increasing sharing, liking, and commenting behaviors, thereby improving engagement [1,26]. On the other hand, videos with longer titles generate greater levels of sharing, liking, and commenting, simultaneously. This is consistent with the findings

of Halvey and Keane [23], who established that YouTube videos with longer titles receive more views. Longer video titles help to enhance the logical and interactive elements of expression but can also deliver more information. In contrast, shorter titles may compromise the integrity and accuracy of video content cues. Hence, the longer the video title is, the more accurately and effectively the core content can be presented, which increases the likelihood of receiving likes, shares, and comments. However, this conflicts with the result of a recent study on YouTube marketing, in which video titles with a lot of information were associated with a lower number of video views [36]. This discrepancy may be due to the evolving functionalities of YouTube, the increasing large volume of available videos, and the changing demographics of YouTube users. With the surge of new videos, the younger generation's attention is precious and they are more likely to be attracted by videos with shorter titles. It remains to be seen whether this change also applies to TikTok.

Second, this study found that content related to the government's handling and guidance for stakeholders promotes citizens' reposting behavior, while information related to the latest news about the crisis reduces the number of likes received. This further expands on the conclusion of Chen et al [1], who found that posts related to the government's handling of information (released by the official Sina Weibo account of the NHCC) promoted citizen engagement on the TikTok platform. This suggests that, whether it is a general social networking platform or one focused on video sharing, the public are concerned about the government's handling of the crisis, which increases the level of citizen engagement. However, compared with Sina Weibo users, content related to the latest news about the COVID-19 crisis has very limited attraction for TikTok users. This may be because short videos that are related to the latest news about the COVID-19 crisis are mostly a combination of plain text and pictures, which are rougher and less interactive.

Interestingly, this study discovered that higher levels of dialogic loop significantly reduce the number of shares received. This conflicts with existing studies that examine how dialogic loops affect citizen engagement [1,25,37]. It is worth noting that previous studies that support the positive impact of dialogic loop on citizen engagement were all concentrated on nonvideo social media platforms, such as Facebook, Twitter, and Sina Weibo, while this study focused on TikTok, a short video-sharing platform. This indicates that the effect of dialogic loop on citizen engagement may vary depending on the type of social media platform. This study measured dialogic loop by video title. Although a video's title is a concentrated expression of the content displayed, it is not completely equivalent to the interactivity of the video itself. In addition, this inconsistency may also be due to different measurements of dialogic loop. In fact, researchers have not reached a consensus on how to evaluate dialogic loop on social media platforms [1]. For example, Men et al [37] measured the dialogic loop of chief executive officers' Facebook accounts from three dimensions, namely replying to comments, liking user comments, and launching surveys. Chen et al [1] adopted 5 dimensions, including using hashtags and mentions, launching surveys, asking users questions, and answering users' questions, to assess the dialogic loop of government social media. Both studies found that dialogic loop is positively associated with social media engagement. A study by Wang and Yang [35] used questions and answers published by nonprofit and for-profit organizations on Twitter to measure dialogic loop. Interestingly, our findings are comparable in that responding to questions resulted in the lowest numbers of likes and retweets, while asking questions had little effect on the number of likes and shares received [35].

Most importantly, this study confirmed that emotional valence can moderate the impact of video length and content type on citizen engagement, although the moderating effects vary in the three dimensions of citizen engagement. For longer videos, the more negative emotion the video's title displays, the lower the number of likes that the video would receive. This pattern is almost the same for the number of comments. This means that longer videos with positive titles receive higher numbers of likes and comments. For short videos related to the government's handling of the pandemic and guidance for stakeholders, positive titles received more shares. Videos about the latest news received more likes if the title had a higher level of positive emotion. This demonstrates the important role of the emotional valence of video titles in promoting citizen engagement through government TikTok accounts. This can be explained by The Social Sharing of Emotion Theory (SSET), which was proposed by Rimé and colleagues in the 1990s. The SSET posits that emotion is a critical motivator for information sharing and social interaction. In our daily life, emotional events can happen anywhere and at any time, and those who experience emotional events are likely to talk about them and the feelings experienced [38]. Through this social sharing of emotion, individuals can express their emotional state and facilitate the establishment of interpersonal relationships [38]. Considering the existence of positivity bias, information with positive emotion creates a good experience for the receiver, and builds a positive image of the information sender [39]. Compared to negative emotion, positive emotion can activate individuals' desire to seek and share information [40]. Therefore, when individuals encounter emotional content from the title of a TikTok video, they will have an urgent need to share this emotional experience. Further, positive emotion will strengthen this behavior, facilitating public engagement on TikTok, enhancing likes, shares, and comments.

Contribution and Implications

This research makes the following contributions. First, we not only empirically examined how governments have used new video-based social media platforms for public engagement in the context of a public health crisis, but also built a model with emotional valence as a moderator to explain the underlying mechanism of how video length, title length, dialogic loop, and content type affects public engagement with government TikTok accounts. Second, we reveal that although shorter videos are likely to attract more likes and comments, for longer videos, the emotional valence of the title is of critical importance. For long videos, the more positive the title, the more likes and comments it receives, and vice versa. Third, our study found that emotional valence displayed in the title plays a critical role in the impact of content type on public engagement. For videos

focused on guidance and the government's handling of the crisis, the more positive the emotion displayed in the video title, the higher the number of shares it received. In addition, videos related to the latest news received more likes when the title had higher levels of positive emotion. Fourth, our study found that dialogic loop is likely to reduce certain types of public engagement.

This study has several practical implications. First, TikTok has become an indispensable channel for government-citizen communication, and governments should actively embrace this new method of communication. Second, for government entities producing videos for TikTok, shorter video length is preferred. If the video length is relatively long, it is better to have a title with positive emotion rather than negative emotion. Third, after the public health crisis, governments are advised to publish more information about their handling of the crisis, provide further guidance for stakeholders, and upload videos with positive emotion in the title to facilitate delivery.

Limitations and Future Recommendations

This study has several limitations. First, we focused on the official TikTok account of the NHCC. Whether the results of this research are equally applicable to local health departments is worthy of investigation. Second, it is worth exploring if the conclusions are similar in Western countries. There are notable differences between TikTok and YouTube [41]; although both of them are video-based social networking platforms, TikTok videos are much shorter than those on YouTube, and there are various easy-to-use templates. The typical length of a TikTok video may be just 30-40 seconds, while YouTube videos can be more than 15 minutes long. In addition, TikTok is a mobile-only social media app, while YouTube operates via both a website and mobile app. Their functionalities also differ; for example, YouTube has a dislike button, while TikTok does not. Furthermore, compared with the individualism advocated by Western countries, China emphasizes collectivism. Moreover, the number of comments found by our study could be influenced by the presence of censorship in China. Third, due to the constraints of the content analysis methods, some important variables were not employed in our research model (eg, camera view). Wang [42] found that the camera view of short videos on TikTok significantly affects the audience's immersion and social presence, which are important elements in promoting citizen engagement. We attempted to use content analysis to determine the type of camera view but it was difficult for the encoders to reach consensus and the interreliability was relatively low. Fourth, the significant negative impact of dialogic loop on reposting behavior deserves further investigation, although we infer that this may be due to differences between TikTok and other non-video–related social media platforms. In addition, our study did not obtain information from NHCC regarding their video and post creation process. Future research could use a qualitative approach to find out how the government TikTok operates in terms of content selection, rules, and strategies during a health crisis. Data from TikTok users could also enrich future studies [43,44].

Conclusions

It is evident that fostering engagement through video sharing on TikTok provides public health departments with the ability to promote citizen engagement and accelerate the dissemination of health information. Nonetheless, how to sustain this positive effect is still unclear. Through analysis of 355 COVID-19–related TikTok videos scraped from the official NHCC account, this study addresses this knowledge gap by considering emotional valence as a moderator that determines the effect that video length, title length, dialogic loop, and content type have on the three dimensions of citizen engagement. In general, shorter videos and longer video titles elicit greater citizen engagement. Citizens prefer to view content about the government's handling of the pandemic and guidance for stakeholders, and actively repost these TikTok videos. We confirmed the important role of emotional valence as a moderator and discovered the specific moderating mechanisms. Longer videos with positive titles received higher numbers of likes and comments. For short videos related to government handling and guidance for stakeholders, positive titles received more shares. Videos related to the latest news received more likes when the title had a higher level of positive emotion. Public health departments should create shorter videos with longer titles, with content focused on the government's handling of the crisis and guidance for stakeholders. Moreover, in the future, video producers should consider the emotional valence of titles and align it with the content type and video length.

Acknowledgments

This work was supported by the National Social Science Foundation of China (grant number 17CXW026).

Conflicts of Interest

None declared.

References

1. Chen Q, Min C, Zhang W, Wang G, Ma X, Evans R. Unpacking the black box: How to promote citizen engagement through government social media during the COVID-19 crisis. Comput Human Behav 2020 Sep;110:106380 [FREE Full text] [doi: 10.1016/j.chb.2020.106380] [Medline: 32292239]
2. Rowe G, Frewer LJ. A Typology of Public Engagement Mechanisms. Science, Technology, & Human Values 2016 Aug 18;30(2):251-290. [doi: 10.1177/0162243904271724]

3. French PE. Enhancing the Legitimacy of Local Government Pandemic Influenza Planning through Transparency and Public Engagement. Public Administration Review 2011 Mar;71(02):253-264. [doi: 10.1111/j.1540-6210.2011.02336.x]
4. Guidry JP, Jin Y, Orr CA, Messner M, Meganck S. Ebola on Instagram and Twitter: How health organizations address the health crisis in their social media engagement. Public Relations Review 2017 Sep;43(3):477-486. [doi: 10.1016/j.pubrev.2017.04.009]
5. Wang Y, Hao H, Platt LS. Examining risk and crisis communications of government agencies and stakeholders during early-stages of COVID-19 on Twitter. Comput Human Behav 2021 Jan;114:106568 [FREE Full text] [doi: 10.1016/j.chb.2020.106568] [Medline: 32982038]
6. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public Engagement and Government Responsiveness in the Communications About COVID-19 During the Early Epidemic Stage in China: Infodemiology Study on Social Media Data. J Med Internet Res 2020 May 26;22(5):e18796 [FREE Full text] [doi: 10.2196/18796] [Medline: 32412414]
7. Teichmann L, Nossek S, Bridgman A, Loewen P, Owen T, Ruths D, et al. Public Health Communication and Engagement on Social Media during the COVID-19 Pandemic. OSF Preprints. :1-41 Preprint published online on July 27, 2020 [FREE Full text] [doi: 10.31219/osf.io/7hypj]
8. Ngai C, Singh R, Lu W, Koon A. Grappling With the COVID-19 Health Crisis: Content Analysis of Communication Strategies and Their Effects on Public Engagement on Social Media. J Med Internet Res 2020 Aug 24;22(8):e21360 [FREE Full text] [doi: 10.2196/21360] [Medline: 32750013]
9. Lovari A, Bowen S. Social media in disaster communication: A case study of strategies, barriers, and ethical implications. J Public Affairs 2019 Jun 13;20(1):A. [doi: 10.1002/pa.1967]
10. Silva P, Tavares AF, Silva T, Lameiras M. The good, the bad and the ugly: Three faces of social media usage by local governments. Government Information Quarterly 2019 Jul;36(3):469-479. [doi: 10.1016/j.giq.2019.05.006]
11. Kim S, Liu BF. Are All Crises Opportunities? A Comparison of How Corporate and Government Organizations Responded to the 2009 Flu Pandemic. Journal of Public Relations Research 2012 Jan;24(1):69-85. [doi: 10.1080/1062726x.2012.626136]
12. Hagen L, Scharf R, Neely S, Keller T. Government social media communications during zika health crisis. : ACM Press; 2018 Presented at: 19th Annual International Conference on Digital Government Research; May 2018; Delft, The Netherlands p. 1-10. [doi: 10.1145/3209281.3209364]
13. Graham MW, Avery EJ, Park S. The role of social media in local government crisis communications. Public Relations Review 2015 Sep;41(3):386-394. [doi: 10.1016/j.pubrev.2015.02.001]
14. Chen S, Xu Q, Buchenberger J, Bagavathi A, Fair G, Shaikh S, et al. Dynamics of Health Agency Response and Public Engagement in Public Health Emergency: A Case Study of CDC Tweeting Patterns During the 2016 Zika Epidemic. JMIR Public Health Surveill 2018 Nov 22;4(4):e10827 [FREE Full text] [doi: 10.2196/10827] [Medline: 30467106]
15. Digital 2019 Q2 Global Digital Statshot. Wearesocial. 2019. URL: http://wearesocial.cn/blog/2019/04/28/digital-2019-q2-global-digital-statshot/ [accessed 2020-05-10]
16. Basch CH, Hillyer GC, Jaime C. COVID-19 on TikTok: harnessing an emerging social media platform to convey important public health messages. Int J Adolesc Med Health 2020 Aug 10:1-3. [doi: 10.1515/ijamh-2020-0111] [Medline: 32776899]
17. The 46th China Statistical Report on Internet Development. CNNIC. 2020. URL: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202009/P020200929546215182514.pdf [accessed 2020-12-20]
18. Ji YG, Chen ZF, Tao W, Cathy Li Z. Public Relations Review 2019 Mar;45(1):88-103. [doi: 10.1016/j.pubrev.2018.12.001]
19. Zhu C, Xu X, Zhang W, Chen J, Evans R. How Health Communication via Tik Tok Makes a Difference: A Content Analysis of Tik Tok Accounts Run by Chinese Provincial Health Committees. Int J Environ Res Public Health 2019 Dec 27;17(1):192 [FREE Full text] [doi: 10.3390/ijerph17010192] [Medline: 31892122]
20. Guo P, Kim J, Rubin R. How video production affects student engagement: an empirical study of MOOC videos. : ACM Press; 2014 Presented at: Proceedings of the first ACM conference on Learning @ scale; May 2014; Atlanta, GA, USA p. 41-50. [doi: 10.1145/2556325.2566239]
21. Lee CH, Yu H. The Impact of Language on Retweeting during Acute Crises: Uncertainty Reduction and Language Expectancy Perspectives. SSRN Journal 2019:1-29. [doi: 10.2139/ssrn.3446718]
22. Xu W, Zhang C. Sentiment, richness, authority, and relevance model of information sharing during social Crises—the case of #MH370 tweets. Computers in Human Behavior 2018 Dec;89:199-206. [doi: 10.1016/j.chb.2018.07.041]
23. Halvey M, Keane M. Analysis of online video search and sharing. : ACM Press; 2007 Presented at: Proceedings of the 18th Conference on Hypertext and Hypermedia; 2007; Manchester, UK p. 217-226. [doi: 10.1145/1286240.1286301]
24. Madichie NO, Hinson R. A critical analysis of the “dialogic communications” potential of sub-Saharan African Police Service websites. Public Relations Review 2014 Jun;40(2):338-350. [doi: 10.1016/j.pubrev.2013.10.009]
25. Gálvez-Rodríguez MDM, Sáez-Martín A, García-Tabuyo M, Caba-Pérez C. Exploring dialogic strategies in social media for fostering citizens’ interactions with Latin American local governments. Public Relations Review 2018 Jun;44(2):265-276. [doi: 10.1016/j.pubrev.2018.03.003]
26. Katz E, Blumler JG, Gurevitch M. Uses and Gratifications Research. The Public Opinion Quarterly 1973:509-523 [FREE Full text] [doi: 10.1086/268109]

27. Wang J, Zhou Y, Zhang W, Evans R, Zhu C. Concerns Expressed by Chinese Social Media Users During the COVID-19 Pandemic: Content Analysis of Sina Weibo Microblogging Data. J Med Internet Res 2020 Nov 26;22(11):e22152 [FREE Full text] [doi: 10.2196/22152] [Medline: 33151894]
28. Thackeray R, Neiger BL, Burton SH, Thackeray CR. Analysis of the purpose of state health departments' tweets: information sharing, engagement, and action. J Med Internet Res 2013 Nov 11;15(11):e255 [FREE Full text] [doi: 10.2196/jmir.3002] [Medline: 24217361]
29. Park H, Reber BH, Chon M. Tweeting as Health Communication: Health Organizations' Use of Twitter for Health Promotion and Public Engagement. J Health Commun 2016 Dec 30;21(2):188-198. [doi: 10.1080/10810730.2015.1058435] [Medline: 26716546]
30. Zhang L, Wei J, Boncella RJ. Emotional communication analysis of emergency microblog based on the evolution life cycle of public opinion. IDD 2020 Feb 27;48(3):151-163. [doi: 10.1108/idd-10-2019-0074]
31. Keib K, Himelboim I, Han J. Important tweets matter: Predicting retweets in the #BlackLivesMatter talk on Twitter. Computers in Human Behavior 2018 Aug;85:106-115. [doi: 10.1016/j.chb.2018.03.025]
32. Araujo T, Kollat J. Communicating effectively about CSR on Twitter. INTR 2018 Apr 04;28(2):419-431. [doi: 10.1108/intr-04-2017-0172]
33. Zhou J, Liu F, Zhou H. Understanding health food messages on Twitter for health literacy promotion. Perspect Public Health 2018 May 07;138(3):173-179. [doi: 10.1177/1757913918760359] [Medline: 29513075]
34. Tang X, Li S, Gu N, Tan M. Exploring repost features of police-generated microblogs through topic and sentiment analysis. EL 2019 Aug 05;37(4):607-623. [doi: 10.1108/el-02-2019-0044]
35. Wang Y, Yang Y. Dialogic communication on social media: How organizations use Twitter to build dialogic relationships with their publics. Computers in Human Behavior 2020 Mar;104:106183. [doi: 10.1016/j.chb.2019.106183]
36. Tafesse W. YouTube marketing: how marketers' video optimization practices influence video views. INTR 2020 Jul 03;30(6):1689-1707. [doi: 10.1108/intr-10-2019-0406]
37. Men LR, Tsai WS, Chen ZF, Ji YG. Social presence and digital dialogic communication: engagement lessons from top social CEOs. Journal of Public Relations Research 2018 Jul 30;30(3):83-99. [doi: 10.1080/1062726x.2018.1498341]
38. Luminet O, Bouts P, Delie F, Manstead ASR, Rimé B. Social sharing of emotion following exposure to a negatively valenced situation. Cognition & Emotion 2000 Sep;14(5):661-688. [doi: 10.1080/026999300501176666]
39. Yang Q, Tufts C, Ungar L, Guntuku S, Merchant R. To Retweet or Not to Retweet: Understanding What Features of Cardiovascular Tweets Influence Their Retransmission. J Health Commun 2018 Nov 07;23(12):1026-1035 [FREE Full text] [doi: 10.1080/10810730.2018.1540671] [Medline: 30404564]
40. Zavattaro SM, French PE, Mohanty SD. A sentiment analysis of U.S. local government tweets: The connection between tone and citizen involvement. Government Information Quarterly 2015 Jul;32(3):333-341. [doi: 10.1016/j.giq.2015.03.003]
41. Che X, Ip B, Lin L. A Survey of Current YouTube Video Characteristics. IEEE MultiMedia 2015 Apr;22(2):56-63. [doi: 10.1109/mmul.2015.34]
42. Wang Y. Humor and camera view on mobile short-form video apps influence user experience and technology-adoption intent, an example of TikTok (DouYin). Computers in Human Behavior 2020 Sep;110:106373. [doi: 10.1016/j.chb.2020.106373]
43. Ahmad AR, Murad HR. The Impact of Social Media on Panic During the COVID-19 Pandemic in Iraqi Kurdistan: Online Questionnaire Study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
44. Yang F, Zhao S, Li W, Evans R, Zhang W. Understanding user satisfaction with Chinese government social media platforms. IR 2020:1. [doi: 10.47989/irpaper865]

Abbreviations

CNNIC: China Internet Network Information Center
IRR: Incidence Rate Ratio
NHCC: National Health Commission of China
SSET: Social Sharing of Emotion Theory

Edited by G Eysenbach, R Kukafka; submitted 16.06.20; peer-reviewed by O Windisch, A Alasmari, J Li, M Adly, A Adly, A Adly, MDG Pimentel, A Ramazan; comments to author 10.09.20; revised version received 15.10.20; accepted 19.01.21; published 04.02.21.

Please cite as:
Chen Q, Min C, Zhang W, Ma X, Evans R
Factors Driving Citizen Engagement With Government TikTok Accounts During the COVID-19 Pandemic: Model Development and Analysis
J Med Internet Res 2021;23(2):e21463
URL: http://www.jmir.org/2021/2/e21463/
doi:10.2196/21463
PMID:33481756

Original Paper

The US Public's Perception of the Threat of COVID-19 During the Rapid Spread of the COVID-19 Outbreak: Cross-Sectional Survey Study

Xiaolei Xiu[1*], MSc; Anran Wang[1*], MSc; Qing Qian[1], MSc; Sizhu Wu[1], PhD

Department of Medical Data Sharing, Institute of Medical Information & Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Sizhu Wu, PhD
Department of Medical Data Sharing
Institute of Medical Information & Library
Chinese Academy of Medical Sciences & Peking Union Medical College
3 Yabao Road
Chaoyang District
Beijing, 100020
China
Phone: 86 10 5232 8760
Email: wu.sizhu@imicams.ac.cn

Abstract

Background: The rapid spread of the COVID-19 pandemic in the United States has made people uncertain about their perceptions of the threat of COVID-19 and COVID-19 response measures. To mount an effective response to this epidemic, it is necessary to understand the public's perceptions, behaviors, and attitudes.

Objective: We aimed to test the hypothesis that people's perceptions of the threat of COVID-19 influence their attitudes and behaviors.

Methods: This study used an open dataset of web-based questionnaires about COVID-19. The questionnaires were provided by Nexoid United Kingdom. We selected the results of a questionnaire on COVID-19–related behaviors, attitudes, and perceptions among the US public. The questionnaire was conducted from March 29 to April 20, 2020. A total of 24,547 people who lived in the United States took part in the survey.

Results: In this study, the average self-assessed probability of contracting COVID-19 was 33.2%, and 49.9% (12,244/24,547) of the respondents thought that their chances of contracting COVID-19 were less than 30%. The self-assessed probability of contracting COVID-19 among women was 1.35 times that of males. A 5% increase in perceived infection risk was significantly associated with being 1.02 times (OR 1.02, 95% CI 1.02-1.02; P<.001) more likely to report having close contact with >10 people, and being 1.01 times (OR 1.01, 95% CI 1.01-1.01; P<.001) more likely to report that cohabitants disagreed with taking steps to reduce the risk of contracting COVID-19. However, there was no significant association between participants who lived with more than 5 cohabitants or less than 5 cohabitants (P=.85). Generally, participants who lived in states with 1001-10,000 COVID-19 cases, were aged 20-40 years, were obese, smoked, drank alcohol, never used drugs, and had no underlying medical conditions were more likely to be in close contact with >10 people. Most participants (21,017/24,547, 85.6%) agreed with washing their hands and maintaining social distancing, but only 20.2% (4958/24,547) of participants often wore masks. Additionally, male participants and participants aged <20 years typically disagreed with washing their hands, maintaining social distancing, and wearing masks.

Conclusions: This survey is the first attempt to describe the determinants of the US public's perception of the threat of COVID-19 on a large scale. The self-assessed probability of contracting COVID-19 differed significantly based on the respondents' genders, states of residence, ages, body mass indices, smoking habits, alcohol consumption habits, drug use habits, underlying medical conditions, environments, and behaviors. These findings can be used as references by public health policy makers and health care workers who want to identify populations that need to be educated on COVID-19 prevention and health.

(J Med Internet Res 2021;23(2):e23400) doi:10.2196/23400

KEYWORDS

COVID-19; perceived threat of disease; environment; behavior; America

Introduction

Background

COVID-19 is an acute infectious respiratory disease that is caused by the novel SARS-CoV-2 virus [1]. It was first detected on December 2019 in Wuhan, China. SARS-CoV-2 is a novel coronavirus that has not been previously identified in humans. It also has a stronger ability to spread and a longer incubation period than the severe acute respiratory syndrome coronavirus. Therefore, SARS-CoV-2 poses a greater epidemic risk [2]. In January 29, 2020, 7711 COVID-19 cases were confirmed in China, and 98 cases were confirmed in 18 other countries [3,4]. As a result, the World Health Organization (WHO) declared the COVID-19 epidemic a public health emergency of international concern on January 30, 2020, and called on all countries to collaborate and prevent the spread of the COVID-19 epidemic [5]. At the beginning of the outbreak, the Chinese government immediately initiated rigorous and intensive prevention and control measures to contain the spread of SARS-CoV-2. Such measures included implementing effective medical treatments, initiating precautionary measures, conducting comprehensive testing, and reducing social mobility and social contact by implementing travel restrictions, banning large gatherings, closing public places, and issuing a stay-at-home order [6,7]. However, due to the strong human-to-human transmissibility of the SARS-CoV-2 virus, the COVID-19 epidemic spread rapidly around the world. On March 11, 2020, the WHO declared the COVID-19 epidemic a worldwide pandemic.

The United States announced the first confirmed case of COVID-19 on January 21, 2020, and the SARS-CoV-2 virus spread to all 50 states by mid-March. As of April 28, 2020, the cumulative number of confirmed cases in the United States has exceeded 1,000,000 [8]. To control the development of the COVID-19 epidemic, the US government implemented a number of measures, including enhancing detection capacities, increasing investments in medical resources, limiting social activities, and providing economic assistance. However, the number of confirmed cases is still rapidly growing at a rate of more than 20,000 new cases per day. The continuous development of the COVID-19 epidemic has seriously affected the physical and mental health of the US public. In terms of coping with large-scale epidemics, the effects of policy implementation are usually influenced by group behaviors. People's methods for coping with large-scale epidemics depend on each individual's perception of the risk of contracting the disease and each individual's ability to adjust their behavior when adapting to environmental changes, to a certain extent [9]. To ensure ultimate success in the fight against COVID-19, it is necessary for the public to adhere to the control measures that have been proposed by the government. However, compliance with control measures depends on the perceived threat level of COVID-19. Therefore, to control the pandemic and quickly reduce the impact of the epidemic on the US economy and society, we must understand the US public's perception of the threat of COVID-19. This will provide the health sector with useful information on increasing disaster preparedness and improving countermeasures.

Risk Perception During the COVID-19 Epidemic

Risk perceptions refer to people's intuitive evaluations of hazards that they are or might be exposed to [10]. Risk perception is influenced by multiple individual and societal factors, such as different social, cultural, and contextual factors. Risk perceptions act as triggers for precautionary action [11]. When individuals and communities deem risks as unsafe, unacceptable, or something to be feared, their responses and adherence to important public health measures for risk mitigation will be influenced [12]. For example, risk perceptions may shape the public's willingness to accept and adhere to COVID-19 risk mitigation strategies, such as social distancing and the use of face masks [13]. As the number of COVID-19–related deaths rises around the world, understanding the public's risk perceptions becomes increasingly more important [14].

Researchers around the world have been actively conducting COVID-19 surveys. Research on the perception of COVID-19 risk mainly focuses on the following 3 aspects: (1) the influencing factors of risk perception (eg, personality characteristics, knowledge level, and health and economy status) and their relationship with protective behaviors [15-17]; (2) the risk perceptions of COVID-19 among special populations, such as college students under quarantine, medical students, patients with COVID-19, health care professionals, and pregnant women [18-22]; and (3) the impact of risk perception on social distancing and the mental health of the public or health workers (eg, anxiety and distress) [23-27]. However, there are few studies on the American public's risk perception of COVID-19. For instance, Wise et al [28] investigated risk perceptions and self-reported protective behaviors among 1591 individuals who lived in the United States during the first week of the outbreak. Furthermore, Niepel et al [29] discussed the risk perception of COVID-19–related fatality among adult US residents. However, there is limited scientific evidence for personality characteristics, risk perceptions, attitudes, and behavior patterns among the American public, especially with regard to the attitudes and behaviors of cohabitants. In addition, most research on the United States is based on adults; there is a lack of research on the risk perceptions of COVID-19 among American people aged <18 years. Moreover, the sample sizes of existing surveys have been limited (ie, generally hundreds or thousands of samples), which has led to certain limitations with regard to the interpretation of the results. Additionally, there have been no COVID-19 surveys that target American people of all ages and have a very large sample size (ie, tens of thousands of samples).

This paper describes a cross-sectional web-based survey that was designed to gauge the US public's perception of the threat posed by COVID-19. This study is one of the first attempts to conduct a large-scale survey for determining the US public's perception of the health threat of COVID-19, and the characteristics that influence the US public's environment and behaviors toward COVID-19.

Methods

Participants

The dataset for this cross-sectional study came from the COVID-19 (Coronavirus) Survival Calculator project [30]. This project was created and managed by the data research team at Nexoid United Kingdom, which is a software company that is located in London, United Kingdom and specializes in research, analysis, and computer science. The COVID-19 (Coronavirus) Survival Calculator is a website that was posted on Reddit (ie, a discussion-based platform) under the subreddit r/Coronavirus_PH. The website was launched to measure participants' risk rates of SARS-CoV-2 infection [30]. The website is available in 49 languages, such as English, Spanish, Portuguese, Russian, and French, and it asks participants certain questions in a survey-like manner (ie, participants can choose the following 3 answers: disagree, neutral, or agree). The dataset is protected by the Creative Commons Attribution 4.0 International (CC BY 4.0) License. The dataset is available for everyone to download, but its respondents remain anonymous for privacy and safety reasons.

On April 20, 2020, Nexoid United Kingdom released the original data of 682,793 surveys. Data were collected from March 24 to April 20, 2020. This open dataset contains the longitude and latitude data of participants' Internet Protocol (IP) addresses. To find respondents who participated in the survey in the United States, we used the Baidu Map geocoder to parse and convert latitude and longitude data, so that we could obtain information on participants' locations. However, to protect participants' privacy, Nexoid United Kingdom does not publish participants' IP addresses. As such, our IP address location data were accurate up to 5 km. Therefore, there may have been several errors in participants' location information in this study.

The dataset initially included 566,122 respondents from the United States. However, we excluded participants who had missing values in the "chance of getting COVID-19" field and participants whose gender was labelled as "other." Furthermore, we conducted stratified random sampling so that the distribution of participants in the age and gender subgroups matched that of the general population, as per the US Census Bureau's methodology when they reported demographic data for the 2019 US census [31]. The respondents participated in this cross-sectional study from March 29 to April 20, 2020. This study was performed 10 weeks after the first COVID-19 case was confirmed in the United States. This survey period corresponds to the rapid spread of the COVID-19 outbreak in the United States.

Data Availability

The data supporting this study is openly available in the Population Health Data Archive [32].

Measures

The COVID-19 (Coronavirus) Survival Calculator is rich in content, such as demographic characteristics, environmental factors, behavioral factors, governmental factors, and health-related factors. In addition, the calculator includes 2 questions about people's perceptions of the threat posed by COVID-19. These questions assessed a respondent's chance of contracting COVID-19 and a respondent's chance of dying from COVID-19. Before participants accessed the survey, Nexoid United Kingdom provided a data and privacy statement, which informed participants that the calculator would record the data that they enter on the page and their locations (ie, locations derived from participants' IP addresses). In addition, participants were informed that the survey was anonymous, the collected data would be added to a freely accessible dataset, and the dataset would be shared.

For the purposes of this study, we selected specific data from the COVID-19 (Coronavirus) Survival Calculator, such as demographics, environmental factors, behavioral factors, and the people's chances of contracting COVID-19. Demographic characteristics included gender, the place of current residence, age, body mass index (BMI), smoking status, alcohol consumption status, nonprescription/recreational drug use (ie, cannabis, amphetamines, cocaine, lysergic acid diethylamide, and 3,4-methylenedioxymethamphetamine use), and underlying medical conditions. Several questions had too many answer options in the original questionnaire, which is not conducive to the analysis of practical problems. Therefore, we regrouped the answer options for several questions.

As per the classification criteria of the Centers for Disease Control and Prevention (CDC), participants' areas of residence were divided into the following six categories, based on the number of confirmed COVID-19 cases that were reported by participants' state of residence: 0-1000 cases, 1001-5000 cases, 5001-10,000 cases, 10,001-20,000 cases, 20,001-40,000 cases, and ≥40,001 cases. Participants were also divided into the following four age categories: 0-20 years, 20-40 years, 40-60 years, and >60 years. Furthermore, participants were divided into the following weight categories based on the WHO BMI criteria: underweight (BMI<18.5 kg/m^2), normal weight (BMI=18.5-24.9 kg/m^2), preobesity (BMI=25-29.9 kg/m^2), and obesity (BMI≥30 kg/m^2). In addition, smoking status was divided into four categories (ie, never, quit, vape, and yes), and alcohol consumption status (ie, never, none in last 14 days, and some in last 14 days) and drug use status (ie, never, none in last 28 days, and some in last 28 days) were divided into three categories. Underlying medical conditions included asthma, carcinoma, chronic kidney disease, compromised immune system, coronary heart disease, chronic obstructive lung disease, diabetes, HIV disease, hypertension, or other chronic illness. We categorized participants' underlying medical condition status as either "have" or "none."

The questionnaire in this study contained seven questions (Table 1). Of these 7 questions, 3 pertained to environmental factors, 3 asked about behavioral factors, and 1 addressed people's perceptions of the threat posed by COVID-19. The question that addressed people's perceptions of the threat posed by COVID-19 provided participants with the following 10 answer options: 0%-10%, 10%-20%, 20%-30%, 30%-40%, 40%-50%, 50%-60%, 60%-70%, 70%-80%, 80%-90%, and 90%-100%. The questionnaire used the median of each probability range as participants' final results (eg, 5%, 15%, and 25%).

Table 1. Questionnaire on the environmental and behavioral factors that affect people's perceptions of the risk of contracting COVID-19.

Topic, question, and options	Value
Environment	
How many people were you in close contact with in the last week, including partners, children, work colleagues, customers, patients, etc? (number of close contacts), mean (SD)	6.9 (6.6)
<10 people, n (%)	19,772 (80.5)
>10 people, n (%)	4774 (19.5)
How many people live in your house/apartment? (number of cohabitants), mean (SD)	3.1 (1.6)
<5 people, n (%)	22,741 (92.6)
>5 people, n (%)	1806 (7.4)
Do you travel to work/school?	
Home (I always worked/studied from home), n (%)	1439 (5.9)
Never (I did not go to work/school before), n (%)	5003 (20.4)
Stopped (I have stopped going to work/school), n (%)	10,469 (42.6)
Critical travel (I still go to work; critical job fields include health care, utilities, military, etc), n (%)	5703 (23.2)
Noncritical travel (ie, I still go to work/school), n (%)	1933 (7.9)
Behaviors	
I am taking steps to reduce my risk (eg, social distancing, washing hands, etc)	
Disagree, n (%)	397 (1.6)
Neutral, n (%)	3133 (12.8)
Agree, n (%)	21,017 (85.6)
The people I live with are taking steps to reduce my risk (social distancing, washing hands)	
Disagree, n (%)	740 (3)
Neutral, n (%)	4956 (20.2)
Agree, n (%)	18,851 (76.8)
Do you wear a mask when outside of your house/flat?	
Rarely, n (%)	14,053 (57.2)
Sometimes, n (%)	4981 (20.3)
Usually, n (%)	5513 (22.5)
Opinion of infection (% chance), mean (SD)	33.2 (21.8)
What do you think are your chances of getting COVID-19?	
0%-10%, n (%)	4236 (17.3)
10%-20%, n (%)	3865 (15.7)
20%-30%, n (%)	4143 (16.9)
30%-40%, n (%)	3061 (12.5)
40%-50%, n (%)	3434 (14)
50%-60%, n (%)	3110 (12.7)
60%-70%, n (%)	1195 (5.9)
70%-80%, n (%)	868 (3.5)
80%-90%, n (%)	355 (1.4)
90%-100%, n (%)	280 (1.1)

Statistical Analysis

We described the frequency of participants' demographic characteristics and various environmental and behavioral factors. Independent samples 2-tailed *t* tests, one-way analysis of variance tests, or Chi-square tests were used to compare the differences between groups of categorical variables, as

appropriate. Furthermore, we used Bonferroni-corrected *P* values for multiple comparisons. In our logistic regression analysis, all demographic, environmental, and behavioral variables were used as independent variables, and participants' self-assessments of their chances of contracting COVID-19 were used as the outcome variable. The logistic regression analysis was conducted to identify factors that were associated with participants' perceptions of the threat of contracting COVID-19. We conducted binary logistic regression analyses to identify factors that were associated with environmental factors and behaviors. Factors were selected by using the backward stepwise method. Odds ratios and their 95% confidence intervals were used to quantify the associations among variables, estimated probabilities of infection, environmental factors, and behaviors. Results were considered statistically significant when $P<.05$. All analyses were conducted from May 2020 to June 2020 with SPSS version 23.0 (IBM Corp).

Results

After excluding 97,883 respondents who had missing values in the "chance of getting COVID-19" field, 1742 respondents whose gender was labelled as "other," and 441,950 respondents due to stratified random sampling, the final sample consisted of 24,547 participants. These respondents lived in all 50 states in the United States. Of the 24,547 respondents in the final sample, 12,465 (50.8%) were women, 6100 (24.9%) were aged <20 years, 6677 (27.2%) were aged 20-40 years, 6191 (25.2%) were aged 40-60 years, and 5579 (22.7%) were aged >60 years. Furthermore, the average BMI of the respondents was 29.41 kg/m^2 (SD 7.74 kg/m^2; range 11.6-87.6 kg/m^2), and 39.2% (9617/24,547) of the respondents were obese (BMI≥30 kg/m^2). Additionally, 56.7% (13,919/24,547) of the participants reported that they never smoked, 51.5% (12,632/24,547) reported that they drank alcohol in the last 14 days before taking the survey, and 47.3% (11,621/24,547) reported that they never used drugs. In terms of medical conditions, 61% (14,983/24,547) of participants did not have the underlying medical conditions that were described in this study. Additional demographic information is included in Table 2.

Table 2. Participants' demographic characteristics and self-assessment results for their risk of contracting COVID-19, stratified by demographic variables (N=24,547).

Demographic characteristics	Number of participants, n (%)	Opinion of infection (% chance), mean (SD)	*t* test (df) or *F* test (df)	*P* value
Gender			10.99[a] (24,464.28)	<.001
Female	12465 (50.8)	34.70 (21.44)		
Male	12082 (49.2)	31.65 (22.01)		
Number of COVID-19 cases in a participant's state of current residence			5.83[b] (5)	<.001
0-1000	606 (2.5)	33.07 (22.39)		
1001-5000	5326 (21.7)	33.91 (21.87)		
5001-10,000	3638 (14.8)	32.71 (21.46)		
10,001-20,000	6937 (23.3)	32.96 (21.71)		
20,001-40,000	6834 (27.8)	32.71 (21.63)		
≥40,001	1206 (4.9)	35.80 (22.97)		
Age (years)			163.72[b] (3)	<.001
0-20	6100 (24.9)	31.33 (21.00)		
20-40	6677 (27.2)	36.59 (22.05)		
40-60	6191 (25.2)	35.28 (22.16)		
>60	5579 (22.7)	28.89 (20.92)		
BMI[c]			18.68[b] (3)	<.001
Underweight (BMI<18.5 kg/m^2)	568 (2.3)	31.55 (21.13)		
Normal weight (BMI=18.5-24.9 kg/m^2)	7136 (29.1)	32.15 (21.61)		
Preobesity (BMI=25-29.9 kg/m^2)	7226 (29.4)	32.70 (21.64)		
Obesity (BMI≥30 kg/m^2)	9617 (39.2)	34.46 (21.98)		
Smoking status			10.52[b] (3)	<.001
Never	13919 (56.7)	32.61 (21.53)		
Quit	5374 (21.9)	33.42 (21.92)		
Vape	2189 (8.9)	34.21 (21.46)		
Yes	3065 (12.5)	34.78 (22.72)		
Alcohol consumption status			48.22[b] (2)	<.001
Never	6165 (25.1)	30.95 (21.70)		
None in last 14 days	5750 (23.4)	33.27 (21.87)		
Some in last 14 days	12632 (51.5)	34.27 (21.69)		
Nonprescription/recreational drug use[d]			52.52[b] (2)	<.001
Never	11621 (47.3)	31.86 (21.74)		
None in last 28 days	6491 (26.4)	34.15 (21.67)		
Some in last 28 days	5065 (20.6)	35.33 (21.68)		
Underlying medical conditions			−10.93[a] (24,545)	<.001
None	14983 (61)	31.99 (21.60)		
Have	9564 (39.0)	35.10 (21.92)		

[a]A 2-tailed *t* test value.

[b]An *F* test value.

[c]BMI: body mass index; participants had a mean BMI of 29.41 kg/m^2 (SD 7.74 kg/m^2).

[d]The total number of participants in this category does not equal 24,547 due to missing data.

In terms of environmental factors, the average number of people that the respondents had close contact with in the week before taking the survey was 6.9 people (SD 6.6 people) (Table 1). In addition, 19.5% (4774/24,547) of participants reported that they had close contact with more than 10 people. In our sample, an average of 3.1 people (SD 1.6 people) lived together in 1 house/apartment, and 92.6% (22,741/24,547) of the respondents reported that less than five people lived in their houses/apartments. Furthermore, 42.6% (10,469/24,547) of the participants stopped working and going to school, while 7.9% (1933/24,547) were still in school or engaged in noncritical work.

In terms of behaviors, most participants (21,017/24,547, 85.6%) were willing to take measures to reduce the risk of infection, but 1.6% (397/24,547) of participants reported that they were not willing to take risk reduction measures. Additionally, 76.8% (18,851/24,547) of the respondents stated that their cohabitants were also taking steps to reduce the risk of infection. However, 57.2% (14,053/24,547) of the respondents reported that they rarely or never wore a mask, and only 22.5% (5513/24,547) stated that they often wore masks.

The average self-assessed probability of COVID-19 infection was 33.2% (SD 21.8%; range 5%-95%); 33% (8101/24,547) of the respondents thought their chances of contracting COVID-19 were <20%, but 2.5% (635/24,547) thought they had a >80% chance of contracting COVID-19 (Table 1). Moreover, participants' perceptions of the threat posed by COVID-19 significantly differed across genders (P<.001), places of residence (P<.001), ages (P<.001), BMI categories (P<.001), smoking statuses (P<.001), alcohol consumption statuses (P<.001), nonprescription/recreational drug use statuses (P<.001), and underlying medical condition statuses (P<.001) (Table 2).

The logistic regression analysis results suggested that there were several important relationships between variables (Table 3). Women thought that their chances of contracting COVID-19 was 1.35 times (OR 1.35, 95% CI 1.28-1.41) that of males. The self-assessed probabilities of contracting COVID-19 for participants who lived in states with 5001-10,000, 10,001-20,000, and 20,001-40,000 COVID-19 cases were 22% (OR 0.78, 95% CI 0.70-0.88; OR 0.78, 95% CI 0.70-0.87; OR 0.78, 95% CI 0.70-0.87, respectively) lower than those of participants who lived in states with >40,000 cases. The self-assessed probabilities of contracting COVID-19 for participants aged 0-20 years, 20-40 years, and 40-60 years were 1.32 times (OR 1.32, 95% CI 1.22-1.43), 1.68 times (OR 1.68, 95% CI 1.56-1.81) and 1.55 times (OR 1.55, 95% CI 1.44-1.66) higher than those of participants aged >60 years, respectively. The self-assessed infection probabilities of participants with a normal BMI were 6% (OR 0.94, 95% CI 0.88-0.99) lower than those of participants who were obese. The self-assessed probabilities of contracting COVID-19 for participants who did not smoke or quit smoking were 8% (OR 1.08, 95% CI 1.00-1.16; OR 1.08, 95% CI 1.00-1.18) higher than those of participants who smoked cigarettes. The self-assessed probabilities of participants who did not drink and did not take drugs were 13% (OR 0.87, 95% CI 0.82-0.92) and 23% (OR 0.77, 95% CI 0.72-0.82) lower than those of participants who did drink and take drugs, respectively. Moreover, participants without underlying medical conditions were 32% (OR 0.68, 95% CI 0.65, 0.72) less likely to contract COVID-19 than participants with these conditions, as per their self-assessment results. Furthermore, compared to participants who had close contact with more than 10 people, participants who were in close contact with less than 10 people had a 43% (OR 0.57, 95% CI 0.53, 0.61) lower self-assessed probability of contracting COVID-19. Participants who engaged in critical work assessed that their probability of contracting COVID-19 was 1.50 times (OR 1.50, 95% CI 1.37-1.65) higher than those of participants who engaged in noncritical work. In addition, participants who were neutral about taking steps to reduce their infection risk thought that they were 21% (OR 0.79, 95% CI 0.73, 0.85) less likely to contract COVID-19 compared to those who agreed with taking steps to reduce their infection risk.

The self-assessed probability of contracting COVID-19 for participants who lived with people that did not take steps to reduce their infection risk were 1.56 times (OR 1.56, 95% CI 1.36-1.79) higher than those of participants who lived with people that did take measures for reducing their infection risk. Respondents who rarely wore masks assessed that their probability of contracting COVID-19 was 34% (OR 0.76, 95% CI 0.72-0.81) lower than that of participants who frequently wore masks.

Table 3. Results of the logistic regression analysis of factors that were associated with participants' self-assessed probability of contracting COVID-19.

Categories	B (SE)	OR (95% CI)	*P* value[a]
Demographic characteristics			
Gender			
Female	.30 (.02)	1.35 (1.28-1.41)[a]	<.001
Male	0[b]	1.0 (referent)	N/A[c]
Number of COVID-19 cases in a participant's state of current residence			
0-1000	−.23 (.09)	0.79 (0.67-0.95)	.01
1001-5000	−.16 (.06)	0.85 (0.76-0.95)	.005
5001-10,000	−.24 (.06)	0.78 (0.70-0.88)	<.001
10,001-20,000	−.25 (.06)	0.78 (0.70-0.87)	<.001
20,001-40,000	−.25 (.06)	0.78 (0.70-0.87)	<.001
≥40,001	0[b]	1.0 (referent)	N/A
Age (years)			
0-20	.28 (.04)	1.32 (1.22-1.43)	<.001
20-40	.52 (.04)	1.68 (1.56-1.81)	<.001
40-60	.44 (.04)	1.55 (1.44-1.66)	<.001
>60	0[b]	1.0 (referent)	N/A
BMI[d]			
Underweight (BMI<18.5 kg/m^2)	−.03 (.08)	0.97 (0.83-1.13)	.67
Normal weight (BMI=18.5-24.9 kg/m^2)	−.07 (.03)	0.94 (0.88-0.99)	.02
Preobesity (BMI=25-29.9 kg/m^2)	−.04 (.03)	0.96 (0.91-1.01)	.13
Obesity (BMI≥30 kg/m^2)	0[b]	1.0 (referent)	N/A
Smoking status			
never	.07 (.04)	1.08 (1.00-1.16)	.053
Quit	.08 (.04)	1.08 (1.00-1.18)	.05
Vape	.08 (.05)	1.09 (0.98-1.20)	.11
Yes	0[b]	1.0 (referent)	N/A
Alcohol consumption status			
Never	−.14 (.03)	0.87 (0.82-0.92)	<.001
None in last 14 days	−.12 (.03)	0.89 (0.84-0.94)	<.001
Some in last 14 days	0[b]	1.0 (referent)	N/A
Nonprescription/recreational drugs use status			
Never	−.26 (.03)	0.77 (0.72-0.82	<.001
None in last 28 days	−.11 (.03)	0.90 (0.84-0.96)	.002
Some in last 28 days	0[b]	1.0 (referent)	N/A
Underlying medical conditions			
None	−.38 (.03)	0.68 (0.65-0.72)	<.001
Have	0[b]	1.0 (referent)	N/A
Environment			
Number of close contacts			

Categories	B (SE)	OR (95% CI)	*P* value[a]
<10 people	−.57 (.04)	0.57 (0.53-0.61)	<.001
<10 people	0[b]	1.0 (referent)	N/A
Number of cohabitants			
<5 people	.06 (.04)	1.06 (0.97-1.16)	.17
> 5 people	0[a]	1.0 (referent)	N/A
Going to work/school			
Home	−.12 (.06)	0.89 (0.78-1.01)	.06
Never	.30 (.05)	0.74 (0.67-0.82)	<.001
Stopped	−.04 (.05)	0.96 (0.87-1.05)	.33
Critical travel	.41 (.05)	1.50 (1.37-1.65)	<.001
Noncritical travel	0[b]	1.0 (referent)	N/A
Behaviors			
Taking steps to reduce infection risk			
Disagree	−.52 (.10)	0.60 (0.49-0.72)	<.001
Neutral	−.24 (.04)	0.79 (0.73-0.85)	<.001
Agree	0[b]	1.0 (referent)	N/A
Cohabitants are taking steps to reduce infection risk			
Disagree	.45 (.07)	1.56 (1.36-1.79)	<.001
Neutral	.15 (.03)	1.17 (1.09-1.24)	<.001
Agree	0[b]	1.0 (referent)	N/A
Wearing a mask			
Rarely	−.27 (.03)	0.76 (0.72-0.81)	<.001
Sometimes	.01 (.04)	1.01 (0.94-1.09)	.72
Usually	0[b]	1.0 (referent)	N/A

[a]Bonferroni *P* values are used to correct for multiple comparisons. Statistical significance was set to a level of *P*<.05.

[b]These parameters were set to 0 because they were redundant.

[c]N/A: not appliable.

[d]BMI: body mass index.

The binary logistic regression analysis results revealed several predictors among the environmental factors (Table 4). When participants' self-assessed risks of contracting COVID-19 increased by 5%, they were 1.02 times (OR 1.02, 95% CI 1.02-1.02) more likely to report that they had close contact with more than 10 people, while their odds of reporting a cessation of work/school decreased by 1% (OR 0.99, 95% CI 0.99-0.99). The odds of having close contact with more than 10 people increased by 82% (OR 1.82, 95% CI 1.52-2.18), 89% (OR 1.89, 95% CI 1.57-2.28), 58% (OR 1.58, 95% CI 1.32-1.89), and 45% (OR 1.45, 95% CI 1.21-1.73) for participants who lived in states with 1001-5000, 5001-10,000, 10,001-20,000, and 20,001-40,000 COVID-19 cases, respectively, compared to those who lived in states with >40,001 confirmed COVID-19 cases. The odds of reporting a cessation of work/school decreased by 34% (OR 0.66, 95% CI 0.58-0.75), 32% (OR 0.68, 95% CI 0.59-0.78), 22% (OR 0.78, 95% CI 0.68-0.88), and 26% (OR 0.86, 95% CI 0.74-0.95) for participants who lived in states with 1000-5000, 5001-10,000 cases, 10,001-20,000, and 20,001-40,000 COVID-19 cases, respectively, compared to those who lived in states with >40,001 COVID-19 cases. Compared to participants aged 20-40 years, participants aged 40-60 years were 30% less likely (OR 0.70, 95% CI 0.64-0.76) to have close contact with more than 10 people in the week before taking the survey, and 14% less likely (OR 0.86, 95% CI 0.75-0.99) to live with more than five people. Participants aged <20 years were 1.68 times (OR 1.68, 95% CI 1.56-1.81) more likely to report that they ceased to go to work/school compared to those aged 20-40 years. Participants who were obese were 1.28 times more likely (OR 1.28, 95% CI 1.18-1.39) to have close contact with more than 10 people compared to participants with a normal BMI. Participants who smoked electronic cigarettes and cigarettes were also 1.41 times (OR 1.41, 95% CI 1.26-1.58) and 1.68 times (OR 1.68, 95% CI 1.52-1.86) more likely to have close contact with more than 10 people compared to nonsmokers, respectively. Participants who

drank alcohol within 14 days before taking the survey were 1.30 times more likely (OR 1.30, 95% CI 1.19-1.42) to have close contact with more than 10 people compared to participants who never drank alcohol. However, participants who were obese were 24% less likely (OR 0.76, 95% CI 0.71-0.81) to stop going to work/school compared to participants with a normal BMI, and participants who smoked cigarettes were 42% less likely (OR 0.58, 95% CI 0.53-0.63) to report that they stopped going to work/school compared to nonsmokers. Compared to participants who never used drugs, participants who used drugs in the last 28 days before taking the survey were 21% less likely (OR 0.79, 95% CI 0.72-0.87) to have close contact with more than 10 people. Participants who never used drugs were also 18% more likely (OR 1.18, 95% CI 1.09-1.27) to cease going to school/work compared to participants who did use drugs. Participants who had underlying medical conditions were 16% less likely (OR 0.84, 95% CI 0.79-0.91) to have close contact with more than 10 people and 9% more likely (OR 1.09, 95% CI 1.03-1.15) to stop going to school/work compared to participants who did not have underlying medical conditions.

Table 4. Results of the binary logistic regression analysis of factors that were significantly associated with environmental factors.

Demographic characteristics	Close contact with >10 people		Living with >5 cohabitants		Stopped going to work/school	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Opinion of infection	1.02 (1.02-1.02)[a]	<.001	1.00 (1.00-1.00)	.85	0.99 (0.99-0.99)[a]	<.001
Gender						
Female	0.96 (0.90-1.02)	.21	1.10 (1.00-1.22)	.053	1.11 (1.05-1.17)[a]	<.001
Male	1.0 (referent)	N/A[b]	1.0 (referent)	N/A	1.0 (referent)	N/A
Number of COVID-19 cases in a participant's state of current residence						
0-1000	1.25 (0.94-1.66)	.12	1.13 (0.77-1.66)	.53	0.80 (0.66-0.98)[c]	.03
1001-5000	1.82 (1.52-2.18)[a]	<.001	1.08 (0.84-1.40)	.54	0.66 (0.58-0.75)[a]	<.001
5001-10,000	1.89 (1.57-2.28)[a]	<.001	1.09 (0.84-1.42)	.54	0.68 (0.59-0.78)[a]	<.001
10,001-20,000	1.58 (1.32-1.89)[a]	<.001	1.12 (0.88-1.44)	.36	0.78 (0.68-0.88)[a]	<.001
20,001-40,000	1.45 (1.21-1.73)[a]	<.001	1.07 (0.83-1.37)	.61	0.84 (0.74-0.95)[d]	.005
≥40,001	1.0 (referent)	N/A	1.0 (referent)	N/A	1.0 (referent)	N/A
Age (years)						
0-20	1.05 (0.96-1.15)	.34	2.05 (1.80-2.33)[a]	<.001	1.68 (1.56-1.81)[a]	<.001
20-40	1.0 (referent)	N/A	1.0 (referent)	N/A	1.0 (referent)	N/A
40-60	0.70 (0.64-0.76)[a]	<.001	0.86 (0.75-0.99)[d]	.04	1.03 (0.96-1.10)	.48
>60	0.32 (0.28-0.36)[a]	<.001	0.25 (0.20-0.31)[a]	<.001	0.50 (0.46-0.55)[a]	<.001
BMI[e]						
Underweight (BMI<18.5 kg/m^2)	0.82 (0.65, 1.03)	.09	0.87 (0.65-1.16)	.35	1.13 (0.94-1.35)	.19
Normal weight (BMI=18.5-24.9 kg/m^2)	1.0 (referent)	N/A	1.0 (referent)	N/A	1.0 (referent)	N/A
Preobesity (BMI=25-29.9 kg/m^2)	1.16 (1.06-1.27)[d]	.001	1.01 (0.88-1.15)	.91	0.88 (0.82-0.94)[a]	<.001
Obesity (BMI≥30 kg/m^2)	1.28 (1.18-1.39)[a]	<.001	1.06 (0.94, 1.20)	.32	0.76 (0.71-0.81)[a]	<.001
Smoking status						
Never	1.0 (referent)	N/A	1.0 (referent)	N/A	1.0 (referent)	N/A
Quit	1.06 (0.97-1.17)	.19	1.02 (0.88-1.18)	.80	0.83 (0.62-0.89)[a]	<.001
Vape	1.41 (1.26-1.58)$_a$	<.001	1.13 (0.96-1.33)	.15	0.68 (0.73-0.75)[a]	<.001
Yes	1.68 (1.52-1.86)[a]	<.001	1.37 (1.17-1.60)[a]	<.001	0.58 (0.53-0.63)[a]	<.001
Alcohol consumption status						
Never	1.0 (referent)	N/A	1.0 (referent)	N/A	1.0 (referent)	N/A
None in last 14 days	1.38 (1.25-1.53)[a]	<.001	0.94 (0.82-1.08)	.39	0.95 (0.88-1.03)	.24
Some in last 14 days	1.30 (1.19-1.42)[a]	<.001	0.72 (0.64-0.82)[a]	<.001	1.17 (1.10-1.26)[a]	<.001
Nonprescription/recreational drug use status						
Never	1.0 (referent)	N/A	1.0 (referent)	N/A	1.0 (referent)	N/A
None in last 28 days	0.87 (0.80-0.94)[c]	.001	0.91 (0.80-1.03)	.14	1.16 (1.09-1.24)[a]	<.001
Some in last 28 days	0.79 (0.72-0.87)[a]	<.001	0.90 (0.78-1.04)	.15	1.18 (1.09-1.27)[a]	<.001
Underlying medical conditions						

Demographic characteristics	Close contact with >10 people		Living with >5 cohabitants		Stopped going to work/school	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Have	0.84 (0.79-0.91)[a]	<.001	1.05 (0.95-1.17)	.37	1.09 (1.03-1.15)[d]	.003
None	1.0 (referent)	N/A	1.0 (referent)	N/A	1.0 (referent)	N/A

[a]Significant at a level of *P*<.001.

[b]N/A: not applicable.

[c]Significant at a level of *P*<.05.

[d]Significant at a level of *P*<.01.

[e]BMI: body mass index.

The binary logistic regression analysis results revealed several predictors among behaviors (Table 5). When participants' self-assessed risks of contracting COVID-19 increased by 5%, they were 1.01 times (OR 1.01, 95% CI 1.01-1.01) more likely to report that their cohabitants disagreed with taking steps to reduce the risk of contracting COVID-19, and their odds of reporting that they rarely wore masks decreased by 1% (OR 0.99, 95% CI 0.99-0.99). The odds of women disagreeing with taking measures to reduce their infection risk and rarely wearing masks were 40% (OR 0.60, 95% CI 0.49-2.74) and 25% (OR 0.75, 95% CI 0.71-0.79) lower than those of males, respectively. Compared to those who lived in states with >40,001 confirmed COVID-19 cases, participants from states with 1001-5000 cases and 5001-10,000 cases were 1.98 times (OR 1.98, 95% CI 1.73-2.25) and 1.80 times (OR=1.80, 95% CI: 1.57-2.06) more likely to report that they rarely wore masks, respectively. Compared to participants aged 20-40 years, participants aged <20 years were 2.96 times (OR 2.96, 95% CI 2.27-3.88), 2.04 times (OR 2.04, 95% CI 1.68-2.47) and 1.48 times (OR 1.48, 95% CI 1.36-1.60) more likely to report that they did not take protective measures, their cohabitants did not take protective measures, and they rarely wore masks, respectively. Participants who were obese were 1.14 times (OR 1.14, 95% CI 1.06-1.21) more likely to report that they rarely wore masks compared to participants with a normal BMI. Compared to nonsmokers, participants who smoked cigarettes were 2.22 times (OR 2.22, 95% CI 1.65-2.99), 1.57 times (OR 1.57, 95% CI 1.25-1.96) and 1.11 times (OR 1.11, 95% CI 1.02-1.21) more likely to report that they did not take protective measures, their cohabitants did not take protective measures, and they rarely wore masks, respectively. Participants who drank alcohol were 1.18 times (OR 1.18, 95% CI 1.10-1.27) more likely to report that they rarely wore masks compared to participants who never drank alcohol. Participants who used drugs within 28 days before answering the survey were 1.47 times (OR 1.47, 95% CI 1.20-1.81) more likely to report that their cohabitants did not agree with taking measures to reduce infection risk compared to participants who never used drugs. Additionally, participants who had underlying medical conditions were 1.31 times (OR 1.31, 95% CI 1.12-1.53) more likely to report that their cohabitants did not agree with measures to reduce infection risk compared to participants without these conditions.

Table 5. Results of the binary logistic regression analysis of significant behavioral factors.

Demographic characteristics	Disagree with taking steps to reduce the risk of contracting COVID-19		Cohabitants disagree with taking steps to reduce the risk of contracting COVID-19		Rarely wears a mask	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Opinion of infection	1.00 (0.99-1.00)	.22	1.01 (1.01-1.01)[a]	<.001	0.99 (0.99-1.00)[a]	<.001
Gender						
Female	0.60 (0.49-0.74)[a]	<.001	1.30 (1.12-1.51)[c]	.001	0.75 (0.71-0.79)[a]	<.001
Male	1.0 (Referent)	N/A[b]	1.0 (Referent)	N/A	1.0 (Referent)	N/A
Number of COVID-19 cases in a participant's state of current residence						
0-1000	1.76 (0.83-3.75)	.14	1.20 (0.71-2.05)	.495	1.59 (1.30-1.95)[a]	<.001
1001-5000	1.32 (0.76-2.29)	.33	0.95 (0.66-1.36)	.76	1.98 (1.73-2.25)[a]	<.001
5001-10,000	1.34 (0.76-2.37)	.32	0.82 (0.56-1.20)	.30	1.80 (1.57-2.06)[a]	<.001
10,001-20,000	1.15 (0.66-1.99)	.62	0.99 (0.69-1.41)	.95	1.62 (1.43-1.84)[a]	<.001
20,001-40,000	1.19 (0.69-2.06)	.53	0.87 (0.61-1.24)	.43	1.31 (1.15-1.49)[a]	<.001
≥40,001	1.0 (Referent)	N/A	1.0 (Referent)	N/A	1.0 (Referent)	N/A
Age (years)						
0-20	2.96 (2.27-3.88)[a]	<.001	2.04 (1.68-2.47)[a]	<.001	1.48 (1.36-1.60)[a]	<.001
20-40	1.0 (Referent)	N/A	1.0 (Referent)	N/A	1.0 (Referent)	N/A
40 to 60 years	0.52 (0.35-0.76)[c]	.001	0.46 (0.35-0.59)[a]	<.001	0.68 (0.63-0.73)[a]	<.001
More than 60 years	0.69 (0.47-1.01)	.06	0.56 (0.43-0.72)[a]	<.001	0.42 (0.39-0.46)[a]	<.001
BMI[d]						
Underweight (BMI<18.5 kg/m^2)	1.10 (0.68-1.76)	.70	1.17 (0.81-1.67)	.41	1.07 (0.89-1.29)	.47
Normal weight (BMI=18.5-24.9 kg/m^2)	1.0 (Referent)	N/A	1.0 (Referent)	N/A	1.0 (Referent)	N/A
Preobesity (BMI=25-29.9 kg/m^2)	0.71 (0.54-0.93)[e]	.014	0.73 (0.59-0.89)[c]	.002	1.06 (0.99-1.13)	.12
Obesity (BMI≥30 kg/m^2)	1.03 (0.81-1.31)	.81	0.88 (0.73-1.04)	.13	1.14 (1.06-1.21)[a]	<.001
Smoking status						
Never	1.0 (Referent)	N/A	1.0 (Referent)	N/A	1.0 (Referent)	N/A
Quit	1.31 (0.96-1.78)	.09	1.18 (0.95-1.46)	.14	0.97 (0.91-1.04)	.42
Vape	1.69 (1.25-2.30)[c]	.001	0.94 (0.73-1.22)	.65	1.03 (0.93-1.14)	.54
Yes	2.22 (1.65-2.99)[a]	<.001	1.57 (1.25-1.96)[a]	<.001	1.11 (1.02-1.21)[c]	.019
Alcohol consumption status						
Never	1.0 (Referent)	N/A	1.0 (Referent)	N/A	1.0 (Referent)	N/A
None in last 14days	1.12 (0.83-1.52)	.47	0.98 (0.80-1.21)	.88	1.17 (1.08-1.27)[a]	<.001
Some in last 14 days	1.31 (0.99-1.73)	.056	0.82 (0.67-1.00)[e]	.047	1.18 (1.10-1.27)[a]	<.001
Nonprescription/recreational drug use status						
Never	1.0 (Referent)	N/A	1.0 (Referent)	N/A	1.0 (Referent)	N/A
None in last 28 days	1.03 (0.78-1.36)	.87	1.15 (0.94, 1.41)	.16	0.98 (0.91-1.04)	.46
Some in last 28 days	1.30 (0.99-1.71)	.06	1.47 (1.20-1.81)[a]	<.001	0.82 (0.76-0.89)[a]	<.001
Underlying medical conditions						

Demographic characteristics	Disagree with taking steps to reduce the risk of contracting COVID-19		Cohabitants disagree with taking steps to reduce the risk of contracting COVID-19		Rarely wears a mask	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Have	0.89 (0.71-1.12)	.32	1.31 (1.12-1.53)[c]	.001	0.84 (0.80-0.89)[a]	<.001
None	1.0 (Referent)	N/A	1.0 (Referent)	N/A	1.0 (Referent)	N/A

[a]Significant at a level of *P*<.001.

[b]N/A: not appliable.

[c] Significant at a level of *P*<.01.

[d]BMI: body mass index.

[e]Significant at a level of *P*<.05.

Discussion

Principal Results

This study analyzed the US public's perception of the threat posed by COVID-19 during the rapid spread of the COVID-19 pandemic. Individuals' risk perceptions affect their protective behaviors during the early stages of a pandemic [28,33]. By extending this idea to the COVID-19 epidemic, we hypothesized that when people believe that they have an increased chance of contracting COVID-19, they will be more willing to follow public health recommendations. This study showed that a high self-assessed probability of contracting COVID-19 was related to the following factors: having close contact with more than 10 people, working at a critical job, having cohabitants that disagreed with taking steps to reduce infection risk, and sometimes wearing masks outside the house. In addition, this study also analyzed environmental factors, behaviors, and their associations with people's self-assessed probabilities of contracting COVID-19. We also identified several demographic factors that were associated with environmental factors, behaviors, and people's self-assessed probabilities of contracting COVID-19. These findings can be used as references by public health policy makers and health care workers who want to identify populations that need to be educated on COVID-19 prevention and health education.

The sample's average self-assessed probability of contracting COVID-19 was 33.2%, and 49.9% (12,399/24.547) of the participants believed that their chances of contracting COVID-19 were less than 30%. This indicated that most respondents were optimistic about the COVID-19 pandemic. Moreover, participants' self-assessed probabilities of contracting COVID-19 differed significantly across respondents' genders, participants' states of residence, ages, BMIs, smoking statuses, alcohol consumption statuses, drug use statuses, underlying disease statuses, environments, and COVID-19–related behaviors. Participants who were obese, were aged 20-40 years, consumed alcohol in the last 14 days before taking the survey, took drugs in the last 28 days before taking the survey, and had underlying medical conditions thought that they had a higher chance of contracting COVID-19. These findings are in line with those of other literature about controversial scientific topics [34-36].

In this study, women felt that they had a higher chance of contracting COVID-19 than men. Women's risk perceptions of COVID-19 also affect their protective behaviors. Women pay more attention to self-protection methods, such as agreeing with taking measures to reduce the risk of contracting COVID-19 and wearing masks, compared to men. However, Chen et al [37] found that more men contracted COVID-19 than women. Studies have also shown that the mortality rate of men is higher than that of women [38]. Nonetheless, in this study, male participants were more reluctant to take measures for reducing their infection risk and less willing to wear masks compared to female participants. These environmental factors and behaviors greatly increase the chance of contracting COVID-19.

We also found that young people (ie, aged <20 years) were more reluctant to wear masks and take measures for reducing the risk of contracting COVID-19 compared to older people. Although the vast majority of COVID-19–related deaths occur in older patients and those with underlying health conditions, this does not mean that young people cannot be infected with SARS-CoV-2. Studies have shown that anyone can become severely ill from contracting COVID-19. The SARS-CoV-2 virus has been infecting young people, and this has resulted in young people spreading COVID-19 [39].

In this study, cigarette smokers reported that their chances of contracting COVID-19 were lower than those of nonsmokers. Although there are no peer-reviewed studies that have evaluated the association between the risk of SARS-CoV-2 infection and smoking, available research suggests that smokers are at a high risk of developing severe COVID-19 outcomes and dying [40,41]. Therefore, it may be necessary to strengthen COVID-19–related prevention and health education for men, young people, and smokers, to increase their understanding of COVID-19.

In general, participants who lived in states where the epidemic was more severe thought that they had a high chance of contracting COVID-19. Therefore, they also paid more attention to maintaining social distance and taking protective measures. These participants were more likely to stop going to work/school and often wear masks. It is worth noting that although the outbreak was less severe in states with 1001-5000 confirmed COVID-19 cases, participants in these states paid less attention to maintaining social distance, were the least likely to stop going to school/work, and were the least willing to take protective measures and wear masks. In addition, these participants thought

they were less likely to contract COVID-19 than those who lived in states with more than 40,001 cases. Environmental factors and people's behaviors toward COVID-19 are likely to result in an increased number of confirmed cases in states with 1001-5000 COVID-19 cases. According to the data provided by USA Facts [8], in April 20, 2020, the states with 1001-5000 reported cases were Arkansas, Delaware, Idaho, Iowa, Kansas, Kentucky, Minnesota, Mississippi, Nebraska, Nevada, New Hampshire, New Mexico, Oklahoma, Oregon, South Carolina, South Dakota, Utah, and Wisconsin. A week later, the number of confirmed cases in Arkansas, Delaware, Iowa, Kansas, and Nebraska increased by more than 50%, and the number of cases in Nebraska increased by 122.74%. In addition, the number of COVID-19 cases in Iowa, Mississippi, South Carolina, and Wisconsin exceeded 5000. Therefore, to avoid a resurgence of COVID-19, people who live in low-risk areas cannot ignore daily preventive measures.

Limiting face-to-face contact with others is the best way to reduce the spread of COVID-19. During the COVID-19 outbreak (ie, March 29 to April 20, 2020) 42.6% (10,469/24,547) of participants stopped going to work/school. This was related to the closure of many schools and businesses in the United States on March 2, 2020 [8]. However, 19.5% (4774/24,547) of the participants reported that they had close contact with more than 10 people in the week before taking the survey. This went against the White House's recommendation on March 16, 2020 (ie, this survey was conducted between March 29 and April 20, 2020) [42]. Those who lived in states with 5001-10,000 cases, were aged 20-40 years, were obese, and smoked had a greater chance of reporting that they had close contact with more than 10 people. In the 2 weeks after the lockdown in Hubei Province, China, only 3.6% of people reported that they went to crowded places [43]. It is possible that China's stringent isolation measures and the Chinese government's strong enforcement of these measures motivated Chinese citizens to comply with government orders. Unlike in China, there is still some debate between the federal US government and the state governments about the need to implement stay-at-home orders and similar measures. Stay-at-home orders are predominantly issued by state governments rather than the federal government. On March 19, 2020, California was the first state to issue a stay-at-home order. As of March 29, 2020, 26 states, such as Illinois, New Jersey, New York, and Washington, have issued similar orders [44]. More than half of Americans are required to stay at home to maintain social distance and reduce close contact with each other. Moreover, between March 29 and April 20, 2020, an additional 17 states and Washington, D.C. issued stay-at-home orders. This may have led the population to change their behaviors toward crowds.

Health care workers are among the most vulnerable groups because of their proximity to patients with COVID-19. In our sample, participants who had critical jobs (eg, health care jobs, utilities jobs, military jobs, etc) also thought they had a high probability of contracting COVID-19 (mean 40.81%, SD 23.05%; Multimedia Appendix 1). In spite of this, only 20.2% (1151/5703) of participants with critical jobs reported that they wore masks regularly, and 57.3% (3267/5703) reported that they rarely wore masks. The lack of protective equipment has further increased the risk of infection in people with critical jobs. According to a report from The Hill [45], on March 19, 2020, health care workers were forced to reuse single-use masks and protective equipment due to shortages caused by the COVID-19 epidemic. The US Department of Health and Human Services have reported that there was a serious shortage of new COVID-19 tests for medical staff and a shortage of protective equipment on a large scale. During the COVID-19 pandemic, the strong risk of infection and the shortage of protective equipment have caused great stress, anxiety, fear and other strong emotions in people with critical jobs. Studies have shown that hospital medical staff who are in charge of patients with COVID-19 have a higher incidence of mental symptoms, such as somatization, obsessive-compulsive behavior, anxiety, hostility, and paranoia, than the general public. Furthermore, hospital medical staff have been experiencing obvious psychological, behavioral, and emotional problems [46,47]. Therefore, in addition to increasing the supply of protective equipment, psychological interventions and counseling should be provided to people with critical jobs, to address their psychological needs.

In this study, participants had different attitudes toward social distancing, washing hands, and wearing a mask. We found that 85.6% (21,017/24,547) of participants agreed with taking steps to reduce the risk of contracting COVID-19 (eg, social distancing and washing hands), while only 22.5% (5513/24,547) usually wore masks. In China, only 2% of people have reported that they do not wear masks outside of home [43]. This is not only related to the personality of US citizens, but also to the decisions of the US government. In China, people have been encouraged—and even mandated—to wear masks outside of home. The CDC has encouraged people to wash their hands frequently and avoid close contact with people. However, the CDC has also recommended that healthy people should not use masks, to ensure that masks are available for frontline health care workers [48]. In Western countries, people do not need to wear a mask unless they are sick. The use of masks is still an evolving process. Many studies and practices have shown that wearing masks is important for slowing the spread of SARS-CoV-2 [49-51]. On April 3, 2020, the CDC revised its guidelines for wearing masks, and recommended that people should wear cloth face coverings in public settings where other social distancing measures are difficult to maintain. Since then, the proportion of US residents who wear masks has rapidly increased.

Limitations

This study has several limitations. First, the scope of our investigation was limited. The recruitment of participants was based on their willingness to participate and access to social networking sites. As such, our participants are not representative of the entire American public. Furthermore, participants were not compensated for participating in this survey. This may reflect participants' prior concerns about contracting COVID-19. Therefore, the respondents may have a higher self-assessment risk of contracting COVID-19 compared to that of the general population. However, we conducted stratified random sampling to achieve a distribution of participants that matched the general population in terms of age and sex, thereby reducing selection

bias to a certain extent. Second, people's perceptions of the threat posed by COVID-19 are influenced by many factors, and these factors are constantly changing. The questionnaire in this study asked participants to answer very specific questions that could not comprehensively cover the complex situations surrounding participants' awareness of their chances of getting COVID-19. For example, the content of the Nexoid United Kingdom questionnaire was updated on April 25, 2020; Nexoid United Kingdom added factors such as private health insurance, income, and race. However, these factors are not included in this study.

Conclusions

In summary, this survey is the first large-scale attempt to describe the determinants of the US public's perception of the threat posed by COVID-19. We also describe COVID-19–related environmental factors and behaviors in the United States, and their relationship with the public's perceptions. Our findings suggest that the US public is generally optimistic; they believe that they have a relatively low chance of contracting COVID-19 during the rapid spread of the COVID-19 outbreak. People, especially women, who live in high-risk areas, are obese, smoke, drink, take drugs, and have chronic or immune-related diseases think that they have a high risk of contracting COVID-19. These people are also more anxious. The struggle to control the spread of COVID-19 will be long and drawn out. Considering the likely future resurgence of COVID-19, it is important to consider implementing specific policies and programs for those who are severely affected by the pandemic, to control the epidemic as soon as possible. Due to the limited questionnaire content, it is necessary to conduct more research on other factors (eg, income and race) that affect the US public's assessment of their chances of contracting COVID-19.

Acknowledgments

The authors would like to thank Nexoid United Kingdom for providing the data. This study is supported by CAMS Innovation Fund for Medical Sciences (grant no. 2020-I2M-2-001).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental materials.
[DOCX File , 123 KB - jmir_v23i2e23400_app1.docx]

References

1. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. bioRxiv Preprint posted online on February 11, 2020. [FREE Full text] [doi: 10.1101/2020.02.07.937862]
2. Jiang X, Rayner S, Luo MH. Does SARS-CoV-2 has a longer incubation period than SARS and MERS? J Med Virol 2020 May;92(5):476-478 [FREE Full text] [doi: 10.1002/jmv.25708] [Medline: 32056235]
3. Tracking the epidemic. China CDC Weekly. URL: http://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm [accessed 2020-05-25]
4. WHO Director-General's statement on IHR Emergency Committee on Novel Coronavirus (2019-nCoV). World Health Organization. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-statement-on-ihr-emergency-committee-on-novel-coronavirus-(2019-ncov) [accessed 2020-05-27]
5. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). World Health Organization. URL: https://tinyurl.com/463gxq7l [accessed 2020-05-10]
6. Lau H, Khosrawipour V, Kocbach P, Mikolajczyk A, Schubert J, Bania J, et al. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. J Travel Med 2020 May 18;27(3):taaa037 [FREE Full text] [doi: 10.1093/jtm/taaa037] [Medline: 32181488]
7. Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. Int J Infect Dis 2020 Apr;93:211-216 [FREE Full text] [doi: 10.1016/j.ijid.2020.02.058] [Medline: 32145465]
8. US coronavirus cases and deaths. USA Facts. URL: https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/ [accessed 2020-05-11]
9. Xu J, Peng Z. People at risk of influenza pandemics: The evolution of perception and behavior. PLoS One 2015;10(12):e0144868. [doi: 10.1371/journal.pone.0144868] [Medline: 26658371]
10. Rohrmann B. Risk perception, risk attitude, risk communication, risk management: A conceptual appraisal. 2008 Jun Presented at: The International Emergency Management Society Annual Conference 2008; June 17-19, 2008; Prague, Czech Republic URL: http://www.rohrmannresearch.net/pdfs/rohrmann-rrr.pdf

11. Wiedemann PM, Schütz H. The precautionary principle and risk perception: experimental studies in the EMF area. Environ Health Perspect 2005 Apr;113(4):402-405 [FREE Full text] [doi: 10.1289/ehp.7538] [Medline: 15811829]
12. Slovic P. Perception of risk. Science 1987 Apr 17;236(4799):280-285. [doi: 10.1126/science.3563507] [Medline: 3563507]
13. Malecki K, Keating JA, Safdar N. Crisis communication and public perception of COVID-19 risk in the era of social media. Clin Infect Dis 2020 Jun 16:ciaa758 [FREE Full text] [doi: 10.1093/cid/ciaa758] [Medline: 32544242]
14. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
15. Führer A, Frese T, Karch A, Mau W, Meyer G, Richter M, et al. [COVID-19: Knowledge, risk perception and strategies for handling the pandemic]. Z Evid Fortbild Qual Gesundhwes 2020 Aug;153-154:32-38 [FREE Full text] [doi: 10.1016/j.zefq.2020.06.002] [Medline: 32601024]
16. Shahzad F, Du J, Khan I, Fateh A, Shahbaz M, Abbas A, et al. Perceived threat of COVID-19 contagion and frontline paramedics' agonistic behaviour: Employing a stressor-strain-outcome perspective. Int J Environ Res Public Health 2020 Jul 15;17(14):5102 [FREE Full text] [doi: 10.3390/ijerph17145102] [Medline: 32679748]
17. Abdelrahman M. Personality traits, risk perception, and protective behaviors of Arab residents of Qatar during the COVID-19 pandemic. Int J Ment Health Addict 2020 Jun 22:1-12 [FREE Full text] [doi: 10.1007/s11469-020-00352-7] [Medline: 32837433]
18. Ding Y, Du X, Li Q, Zhang M, Zhang Q, Tan X, et al. Risk perception of coronavirus disease 2019 (COVID-19) and its related factors among college students in China during quarantine. PLoS One 2020;15(8):e0237626. [doi: 10.1371/journal.pone.0237626] [Medline: 32790791]
19. Taghrir MH, Borazjani R, Shiraly R. COVID-19 and Iranian medical students; a survey on their related-knowledge, preventive behaviors and risk perception. Arch Iran Med 2020 Apr 01;23(4):249-254. [doi: 10.34172/aim.2020.06] [Medline: 32271598]
20. Zhong Y, Liu W, Lee TY, Zhao H, Ji J. Risk perception, knowledge, information sources and emotional states among COVID-19 patients in Wuhan, China. Nurs Outlook 2020 Aug 22;69(1):13-21 [FREE Full text] [doi: 10.1016/j.outlook.2020.08.005] [Medline: 32980153]
21. Peres D, Monteiro J, Almeida M, Ladeira R. Risk perception of COVID-19 among the Portuguese healthcare professionals and general population. J Hosp Infect 2020 May 30;105(3):434-437 [FREE Full text] [doi: 10.1016/j.jhin.2020.05.038] [Medline: 32485198]
22. Yue C, Liu C, Wang J, Zhang M, Wu H, Li C, et al. Association between social support and anxiety among pregnant women in the third trimester during the coronavirus disease 2019 (COVID-19) epidemic in Qingdao, China: The mediating effect of risk perception. Int J Soc Psychiatry 2020 Jul 09:20764020941567 [FREE Full text] [doi: 10.1177/0020764020941567] [Medline: 32643510]
23. Ding Y, Xu J, Huang S, Li P, Lu C, Xie S. Risk perception and depression in public health crises: Evidence from the COVID-19 crisis in China. Int J Environ Res Public Health 2020 Aug 07;17(16):5728 [FREE Full text] [doi: 10.3390/ijerph17165728] [Medline: 32784792]
24. Kuang J, Ashraf S, Das U, Bicchieri C. Awareness, risk perception, and stress during the COVID-19 pandemic in communities of Tamil Nadu, India. Int J Environ Res Public Health 2020 Sep 30;17(19):7177 [FREE Full text] [doi: 10.3390/ijerph17197177] [Medline: 33007992]
25. Krok D, Zarzycka B. Risk perception of COVID-19, meaning-based resources and psychological well-being amongst Healthcare Personnel: The mediating role of coping. J Clin Med 2020 Oct 08;9(10):3225 [FREE Full text] [doi: 10.3390/jcm9103225] [Medline: 33050068]
26. Simione L, Gnagnarella C. Differences between health workers and general population in risk perception, behaviors, and psychological distress related to COVID-19 spread in Italy. Front Psychol 2020 Sep 4;11:2166 [FREE Full text] [doi: 10.3389/fpsyg.2020.02166] [Medline: 33013555]
27. Xie K, Liang B, Dulebenets MA, Mei Y. The impact of risk perception on social distancing during the COVID-19 pandemic in China. Int J Environ Res Public Health 2020 Aug 27;17(17):6256 [FREE Full text] [doi: 10.3390/ijerph17176256] [Medline: 32867381]
28. Wise T, Zbozinek TD, Michelini G, Hagan CC, Mobbs D. Changes in risk perception and self-reported protective behaviour during the first week of the COVID-19 pandemic in the United States. R Soc Open Sci 2020 Sep 16;7(9):200742 [FREE Full text] [doi: 10.1098/rsos.200742] [Medline: 33047037]
29. Niepel C, Kranz D, Borgonovi F, Emslander V, Greiff S. The coronavirus (COVID-19) fatality risk perception of US adult residents in March and April 2020. Br J Health Psychol 2020 Nov;25(4):883-888 [FREE Full text] [doi: 10.1111/bjhp.12438] [Medline: 32519364]
30. COVID-19 (Coronavirus) Survival Calculator. Nexoid United Kingdom. URL: https://www.covid19survivalcalculator.com/en/calculator [accessed 2020-05-01]
31. United States population by age and sex. United States Census Bureau. URL: https://www.census.gov/popclock/data_tables.php?component=pyramid [accessed 2021-01-25]

32. Institute of Medical Information, Chinese Academy of Medical Sciences. A cross-sectional survey data set of US public's perception of the threat of COVID-19 during the rapid spread of the COVID-19 outbreak. Population Health Data Archive 2020 May 02 [FREE Full text] [doi: 10.12213/11.A001Z.202101.268.V1.0]
33. Liao Q, Wu P, Lam WWT, Cowling BJ, Fielding R. Trajectories of public psycho-behavioural responses relating to influenza A(H7N9) over the winter of 2014-15 in Hong Kong. Psychol Health 2019 Feb;34(2):162-180. [doi: 10.1080/08870446.2018.1515436] [Medline: 30430862]
34. People with certain medical conditions. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html [accessed 2021-01-27]
35. Ekiz T, Pazarlı AC. Relationship between COVID-19 and obesity. Diabetes Metab Syndr 2020;14(5):761-763 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.047] [Medline: 32505980]
36. Dubey MJ, Ghosh R, Chatterjee S, Biswas P, Chatterjee S, Dubey S. COVID-19 and addiction. Diabetes Metab Syndr 2020;14(5):817-823 [FREE Full text] [doi: 10.1016/j.dsx.2020.06.008] [Medline: 32540735]
37. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
38. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020 Aug;584(7821):430-436. [doi: 10.1038/s41586-020-2521-4] [Medline: 32640463]
39. Bai N. Coronavirus is sickening young adults and spreading through them, experts say. University of California San Francisco. 2020 Mar 24. URL: https://www.ucsf.edu/news/2020/03/416961/coronavirus-sickening-young-adults-and-spreading-through-them-experts-say [accessed 2020-06-04]
40. Berlin I, Thomas D, Le Faou AL, Cornuz J. COVID-19 and smoking. Nicotine Tob Res 2020 Aug 24;22(9):1650-1652 [FREE Full text] [doi: 10.1093/ntr/ntaa059] [Medline: 32242236]
41. Coronavirus disease (COVID-19): Tobacco. World Health Organization. 2020 May 27. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19-tobacco [accessed 2020-11-02]
42. Vaughan A. Covid-19 escalates. New Sci 2020 Mar 21;245(3274):7 [FREE Full text] [doi: 10.1016/S0262-4079(20)30569-8] [Medline: 32518448]
43. Zhong BL, Luo W, Li HM, Zhang QQ, Liu XG, Li WT, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020 Mar 15;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]
44. Lee A. These states have implemented stay-at-home orders. Here's what that means for you. CNN. 04 Jul 07. URL: https://www.cnn.com/2020/03/23/us/coronavirus-which-states-stay-at-home-order-trnd/index.html [accessed 2020-06-07]
45. Johnson M. Health care workers forced to reuse single-use masks, protective equipment as coronavirus causes shortages. The Hill. 2020 Mar 19. URL: https://thehill.com/policy/healthcare/488561-health-care-workers-forced-to-reuse-single-use-masks-protective-equipment [accessed 2020-06-19]
46. Xing J, Sun N, Xu J, Geng S, Li Y. Study of the mental health status of medical personnel dealing with new coronavirus pneumonia. medRxiv Preprint posted online on March 6, 2020. [FREE Full text] [doi: 10.1101/2020.03.04.20030973]
47. Delgado D, Quintana FW, Perez G, Liprandi AS, Ponte-Negretti C, Mendoza I, et al. Personal safety during the COVID-19 pandemic: Realities and perspectives of healthcare workers in Latin America. Int J Environ Res Public Health 2020 Apr 18;17(8):2798 [FREE Full text] [doi: 10.3390/ijerph17082798] [Medline: 32325718]
48. Geggel L. Everyone should wear face 'masks' in public, CDC now recommends. LiveScience. 2020 Apr 03. URL: https://www.livescience.com/cdc-recommends-face-masks-coronavirus.html [accessed 2020-06-13]
49. Choi S, Ki M. Estimating the reproductive number and the outbreak size of COVID-19 in Korea. Epidemiol Health 2020;42:e2020011. [doi: 10.4178/epih.e2020011] [Medline: 32164053]
50. Sun P, Lu X, Xu C, Sun W, Pan B. Understanding of COVID-19 based on current evidence. J Med Virol 2020 Jun;92(6):548-551 [FREE Full text] [doi: 10.1002/jmv.25722] [Medline: 32096567]
51. Offeddu V, Yung CF, Low MSF, Tam CC. Effectiveness of masks and respirators against respiratory infections in healthcare workers: A systematic review and meta-analysis. Clin Infect Dis 2017 Nov 13;65(11):1934-1942 [FREE Full text] [doi: 10.1093/cid/cix681] [Medline: 29140516]

Abbreviations

BMI: body mass index
CDC: Centers for Disease Control and Prevention
IP: Internet Protocol
WHO: World Health Organization

Edited by G Eysenbach; submitted 11.08.20; peer-reviewed by J Xu, D Berrigan; comments to author 31.10.20; revised version received 19.11.20; accepted 19.01.21; published 08.02.21.

Please cite as:
Xiu X, Wang A, Qian Q, Wu S
The US Public's Perception of the Threat of COVID-19 During the Rapid Spread of the COVID-19 Outbreak: Cross-Sectional Survey Study
J Med Internet Res 2021;23(2):e23400
URL: http://www.jmir.org/2021/2/e23400/
doi:10.2196/23400
PMID:33481753

Original Paper

Interdisciplinary Online Hackathons as an Approach to Combat the COVID-19 Pandemic: Case Study

Katarina Braune[1,2,3], MD; Pablo-David Rojas[2], PhD; Joscha Hofferbert[2], MSc; Alvaro Valera Sosa[2,4,5], MSc; Anastasiya Lebedev[2], MSc; Felix Balzer[6,7], MD, PhD, MSc; Sylvia Thun[3], MD; Sascha Lieber[6,8], MD; Valerie Kirchberger[8], MD; Akira-Sebastian Poncette[2,3,6,7], MD

[1]Department of Paediatric Endocrinology and Diabetes, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

[2]Hacking Health Berlin, Berlin, Germany

[3]Berlin Institute of Health, Berlin, Germany

[4]CityLAB Berlin, Building Health Lab, Berlin, Germany

[5]Department of Design and Typologies, Technische Universität Berlin, Berlin, Germany

[6]Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Department of Anesthesiology and Intensive Care Medicine, Berlin, Germany

[7]Einstein Center Digital Future, Berlin, Germany

[8]Executive Board, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

Corresponding Author:
Katarina Braune, MD
Department of Paediatric Endocrinology and Diabetes
Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin
Humboldt-Universität zu Berlin, and Berlin Institute of Health
Augustenburger Platz 1
Berlin, 13353
Germany
Phone: 49 30450616454
Email: katarina.braune@charite.de

Abstract

Background: The COVID-19 outbreak has affected the lives of millions of people by causing a dramatic impact on many health care systems and the global economy. This devastating pandemic has brought together communities across the globe to work on this issue in an unprecedented manner.

Objective: This case study describes the steps and methods employed in the conduction of a remote online health hackathon centered on challenges posed by the COVID-19 pandemic. It aims to deliver a clear implementation road map for other organizations to follow.

Methods: This 4-day hackathon was conducted in April 2020, based on six COVID-19–related challenges defined by frontline clinicians and researchers from various disciplines. An online survey was structured to assess: (1) individual experience satisfaction, (2) level of interprofessional skills exchange, (3) maturity of the projects realized, and (4) overall quality of the event. At the end of the event, participants were invited to take part in an online survey with 17 (+5 optional) items, including multiple-choice and open-ended questions that assessed their experience regarding the remote nature of the event and their individual project, interprofessional skills exchange, and their confidence in working on a digital health project before and after the hackathon. Mentors, who guided the participants through the event, also provided feedback to the organizers through an online survey.

Results: A total of 48 participants and 52 mentors based in 8 different countries participated and developed 14 projects. A total of 75 mentorship video sessions were held. Participants reported increased confidence in starting a digital health venture or a research project after successfully participating in the hackathon, and stated that they were likely to continue working on their projects. Of the participants who provided feedback, 60% (n=18) would not have started their project without this particular hackathon and indicated that the hackathon encouraged and enabled them to progress faster, for example, by building interdisciplinary teams, gaining new insights and feedback provided by their mentors, and creating a functional prototype.

Conclusions: This study provides insights into how online hackathons can contribute to solving the challenges and effects of a pandemic in several regions of the world. The online format fosters team diversity, increases cross-regional collaboration, and can be executed much faster and at lower costs compared to in-person events. Results on preparation, organization, and evaluation of this online hackathon are useful for other institutions and initiatives that are willing to introduce similar event formats in the fight against COVID-19.

(J Med Internet Res 2021;23(2):e25283) doi:10.2196/25283

KEYWORDS

hackathon; COVID-19; digital health; mentoring; interdisciplinarity; interoperability; SARS-CoV-2; public health; innovation; collaboration; hack; mentor; case study; online health care; challenge; implementation; plan; collaboration

Introduction

The COVID-19 outbreak is first and foremost a human tragedy, affecting the health of millions of people and overwhelming health care systems and the global economy [1]. This devastating pandemic has brought together communities across the globe to tackle this issue in an unprecedented manner.

Hackathons are collaborative multiday events that bring people from different professional and personal backgrounds face to face in a venue to work on specific challenges. Previous publications have shown how health hackathons work as an effective model of interdisciplinary collaboration, capable of forwarding recommendations and viable solutions effectively [2-36]; however, they often require several months of preparation and considerable financing [6,34].

Little is so far known about how health hackathons can be executed remotely and in quick response to ongoing circumstances, as well as how they can contribute to solving challenges related to the COVID-19 pandemic, considering the amount and fast pace of evidence gathered on its impact and possible solutions [37-48].

The following case study describes the execution of an online hackathon dedicated to the challenges of the COVID-19 pandemic and participant experience, which will be useful to other organizations and institutions in the health care sector.

Methods

Study Setting

Hacking Health is a global nonprofit organization that pairs innovators with health care experts [6]. Its Berlin chapter organized its first online hackathon in partnership with Charité – Universitätsmedizin Berlin (the largest university hospital in Europe), the Berlin Institute of Health, Diabetes Center Berne, and Data Natives (an online community of data scientists and developers). The event, called EasterHack [49], ran from Friday evening to Monday afternoon over the Easter weekend in 2020. Participants included physicians, health care professionals, researchers, patients, entrepreneurs, engineers, designers, developers, etc. Participation was open to everyone worldwide and was free of charge.

Challenges

The most critical issues caused by, or associated with, COVID-19 and the current state of possible solutions were reviewed and discussed in a series of five virtual meetings held across the interdisciplinary Hacking Health Berlin community in cooperation with the hospital's executive board and different departments of Charité – Universitätsmedizin Berlin.

A summary of this review was sent to, and discussed with, frontline clinicians and researchers for consideration, to ponder on its relevance, and to raise other questions related to the most pressing issues in their line of work. It was highly relevant for identifying unmet needs and urgent problems among health care professionals, vulnerable patient groups, and more generally on a population level, rather than ideas for solutions. Thereafter, a total of six challenges were identified by the Hacking Health Berlin team (Table 1) in close cooperation with the university hospital's executive board and other institutes and departments (Institute of Hygiene and Tropical Medicine, Department of Anesthesiology and Intensive Care Medicine, Department of Psychosomatic Medicine, Department of Paediatric Endocrinology and Diabetes of Charité – Universitätsmedizin Berlin) and the open-source initiative CoEpi.org, an international online community of volunteers with different skillsets working on open-source solutions for privacy-first contact tracing. Participants were allowed to use their choice of technology to solve these challenges, without constraint or limitation.

Table 1. Overview of the hackathon's challenges. A total of six COVID-19–related challenges were identified in close cooperation with frontline physicians and researchers.

Challenge	Questions/considerations
1. Protect high-risk populations	• How can we better support the medical needs of high-risk populations (eg, people with pre-existing conditions such as cancer, diabetes, and immunodeficiencies)? • How can we identify high-risk patients and isolate them from the general population? • How can we support high-risk patients with preventive ambulatory treatment? • How can we deliver care for people with chronic conditions during a pandemic?
2. Protect health care workers	• How can we prevent medical staff from getting infected with SARS-CoV-2 in health care facilities? • How can we manage the increasing demand for personal protective equipment in patient care?
3. Privacy-first contact tracing and digital epidemiology	• What can we build on top of open-source solutions (such as CoEpi.org and COVID-watch.org) to allow users to opt in to disclose anonymous, nonidentifiable information that public health officials would find useful? • How can we authenticate positive COVID-19 tests without requiring public health authorities to do extra work to authenticate them for us? • How can we collect more information on Bluetooth low-energy proximity contact event duration and distance to inform risk scores? • How can we automate the provisional diagnosis of COVID-19 based solely on self-reported symptoms and self-collected data?
4. Improve intensive care	• How can mechanical ventilation be improved to reduce long-term damage to the lungs? • How can we ease prone positioning in ventilated patients? • How can we reduce the formation of aerosols? • How can we visually monitor agitated and hyperactive patients in the ICU[a] to prevent them from extubating or hurting themselves? • How can we give confused ICU patients in-room, time, and situation orientation? • How can we make the ICU a more natural and pleasant environment? • How can we apply evidence-based medicine in all ICUs by tele-ICU? • How can we prevent long-term complications following ICU treatment such as controlled cortical impact and postintensive care syndrome?
5. Protect our mental health	• How can we support health care workers in dealing with the stress and trauma that they might experience during a pandemic? • How can we identify and offer support to people, families, friends, and/or coworkers who are running into mental health issues? • How can we predict critical mental states through early detection?
6. Your idea	• Find a novel solution for your individual COVID-19 challenge, if it has not been listed above

[a]ICU: intensive care unit.

Preparation

The core organizing team consisted of 7 volunteers to set up digital tools; scheduling; marketing outreach strategy for recruitment of participants, cooperating partners, participant and mentor communication; and coordination. An implementation strategy previously designed and used for in-person hackathons by Charité – Universitätsmedizin Berlin and Hacking Health Berlin [6] was adjusted accordingly.

Due to the novelty and urgency of the topic, organization of the hackathon began only 2 weeks prior to the event being completely prepared and executed online, which is different from former hackathons. In April 2020, travel and assembly restrictions, as well as social distancing recommendations, did not allow in-person meetings in Europe. However, the online nature of the event allowed participants and mentors from all over the world to participate regardless of their location, avoiding traveling costs and time.

Participant Engagement

Recent innovation studies have demonstrated interdisciplinarity among participants and within-team compositions are essential for the success of hackathon projects [50,51]. Therefore, it is essential to recruit participants from various relevant fields of expertise such as biomedical research, health care, psychology, education, design, economics, data science, computer science, engineering, social sciences, and politics. Future end-users of the solutions created at a hackathon should be included as early as possible to actively contribute to the cocreation process.

To attract and recruit a diverse and interdisciplinary group of participants, previously existing multidisciplinary networks on LinkedIn [52] and Slack [53] actively engaged in health care innovation, and cooperating with Hacking Health Berlin, were targeted. This provided a great opportunity to reach participants, mentors, and cooperation partners who supported organizers and participating teams in developing projects before, during, and after the hackathon. As one of the largest local networks of innovation in health care, Hacking Health Berlin was able to recruit a large number of interdisciplinary skilled participants

within only 5 days. Professionals working in relevant fields and interested in digital health were also targeted through promotional campaigns using LinkedIn, Facebook, Meetup, and other social media channels.

Participants and mentors were able to sign up via an online form, where they provided information on their skillset, shared their experience level and accomplishments from previous hackathons, their motivations to participate, their desired team role during the hackathon, and a brief outline of their project idea.

Online Tools

To prepare for the hackathon, the organizers used Airtable [54] to arrange schedules, responsibilities, marketing, registration procedures, jury criteria definition, and the collection and structuring of all information provided to the teams before, throughout, and after the event. The Hacking Health Berlin team and external co-organizers had access to this platform.

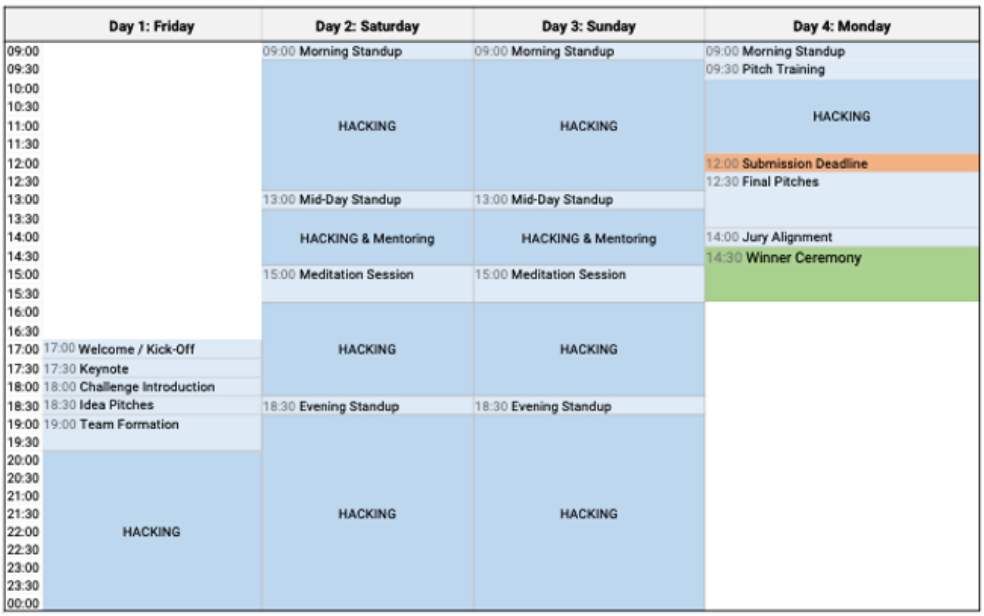
The online instant messenger tool Slack [53] was used for communication during the event to coordinate actions and share data via text and video. Public and closed subchannels were created for the organizing team, mentors, judges, and participants to communicate and execute activities. Additionally, participating teams created their private subchannels on Slack to communicate internally.

Mentoring sessions were coordinated and held on the Mentornity platform [55], as described in a subsequent point. For the opening and closing ceremonies, moderations, pitches, and prototype showcasing sessions, video conferences were held via Zoom and live streamed to the public on Facebook and YouTube.

The Hackathon

The hackathon was conducted over a period of 4 days, from Friday to Monday afternoon. An example schedule is shown in Figure 1.

Figure 1. Activities and schedule of a 4-day online hackathon.



A keynote from the director of the interoperability unit of the Berlin Institute of Health opened the event, motivated the teams, and pointed out the importance of this online hackathon to create solutions for the challenges of the COVID-19 pandemic.

Following the opening ceremony, individual participants and pre-existing teams pitched their ideas for the challenges they selected. Additional participants joined the team of their choice. Team composition is a crucial part of hackathons. As previous research shows, highly diverse and multidisciplinary teams are more likely to be effective in developing creative and sustainable solutions [50,51]. This is particularly relevant in the complex field of digital health, where experts from different fields and areas are needed for development such as health care workers, patients and caregivers, developers, engineers, data scientists, designers, and experts in economics, legal and regulatory affairs. Team building was supported by volunteers from the Hacking Health Berlin team to foster interdisciplinary and international cooperation. A maximum of 6 people per team was recommended.

After teams were formed, the idea and prototype were developed collaboratively over the course of 4 days aided by mentoring sessions. Daily morning and evening stand-up sessions were held virtually, open to all organizers, mentors, and participants.

On day 4, solutions and prototypes were presented to the public and evaluated by an interdisciplinary jury entitled to award the most elaborate and promising solutions based on the criteria described below.

Mentoring

Mentors provided participants with valuable expert knowledge about medicine, design, business, IT (information technology), and other areas of expertise. They indicated their availability for mentoring within the 48-hour span of days 2 and 3. This setting was well embraced, especially by health care

professionals—despite the pressing pandemic situation—and other professionals working from home.

One week prior to the hackathon event, mentors with previous mentoring experience at Hacking Health Berlin events were recruited via email. New mentors were recruited through various local and international networks with the use of social media campaigns and direct messaging via networking platforms. In both cases, mentors were asked to fill in an online form indicating fields of expertise and availability (ie, convenient time slots). This information was transferred to the online platform Mentornity. The platform showed each mentor's profile and facilitated video call sessions for each team. Simultaneously, participants created team profiles for their project, searched mentors by expertise area, and booked at least one mandatory mentoring session made available by the mentor. The platform's analysis feature tracked the number of mentoring sessions held by every mentor and team. Likewise, mentors were encouraged to sign up for Slack—the hackathon's communication platform—to keep everyone informed on sessions prior to and during the event.

Two experts in the field of communication and public speaking offered dedicated pitch training to help teams develop focused final pitches that adhered to the 3-minute time limit.

Jury and Project Evaluation

The interdisciplinary jury was composed of 10 members: a professor of anesthesiology and computer science, a pediatrician and hospital executive board member, a professor of medical informatics, an entrepreneur with a background in open-source technology, a medical doctor and high-risk patient, a senior doctor in hygiene, a digital health connector, a designer, a chief information and digital officer, and the managing director of a research institution.

During the final presentations, teams were asked to first name the challenge or problem intended to be solved, then showcase the solution developed, and how it can be integrated into health care systems. The winning teams were awarded monetary prizes of 5000 EUR (US $6059) in total following the evaluation criteria shown in Table 2.

Table 2. Evaluation criteria for the projects presented in the final pitch session at the end of the hackathon.

Criterion	Considerations
Innovation	How innovative is the idea? Does the approach stand out from solutions that are available so far?
Impact	How high is the benefit for the patient and other stakeholders of the health care system? Does the solution generate impact for our society? How viable is the idea?
Applicability	How user-friendly is the solution? Is the solution convenient to use? How viable is the solution?
Feasibility	How realistic is the implementation of the idea? Can the solution be brought to life by further research projects or by founding a company?
Execution	How sophisticated and well-made is the technical implementation?

Data Collection and Analysis

Two online surveys were designed by the Hacking Health Berlin team, one survey targeting participants and another one for mentors of the hackathon (Multimedia Appendices 1 and 2). The research team members comprised health care professionals, scientists, and entrepreneurs with backgrounds in anesthesiology, intensive care medicine, pediatrics, psychology, microbiology, medical informatics, neuroscience, human factors, health care building design, human-centered design, and marketing. Google Forms was used to collect responses. CHERRIES (Checklist for Reporting Results of Internet E-Surveys) was used to guide survey development (Multimedia Appendix 3) [56].

The participant survey comprised 17 plus 5 optional items (using branching logic, dependent on responses to previous questions). The questions assessed quality experience throughout the online hackathon, including the mentorship sessions and future plans regarding the projects presented. Furthermore, mentors provided feedback to the organizers via an online survey through Mentornity subsequently to each session (Multimedia Appendix 1). All data were collected during a single event with surveys being distributed to mentors and participants at the end of the hackathon. Surveys were anonymous and voluntary and included electronic consent. Participant responses were collected, managed, and analyzed using Google Forms. A thematic hackathon T-shirt was offered as a reward for completing the surveys.

Results

Participants

In total, 171 participants, of whom 42% (n=72) were female, expressed their interest and preregistered online within 5 days prior to the hackathon. Of the total, 84 participants had joined the Zoom event for the opening ceremony, 62 of whom formed 23 different teams. Finally, 14 teams comprising 48 members pitched their completed projects at the end of the hackathon.

Participants based in at least 8 countries participated, including Germany (n=25, 52%), the United States (n=5, 10%), Bosnia and Herzegovina (n=3, 6%), Canada (n=3, 6%), and others (n=5, 10%) like Belgium, Sri Lanka, Switzerland, and the United Kingdom, as well as 7 participants (15%) from unknown locations. Various professional backgrounds were represented, with IT (n=8, 17%), research (n=5, 10%), project management (n=4, 8%), and software engineering (n=4, 8%) being the most frequent work fields. Other backgrounds (n=27, 56%) included business development, data science, entrepreneurship, health care, human resource management, mechanical engineering, quality management, education, and UI/UX (user interface/user

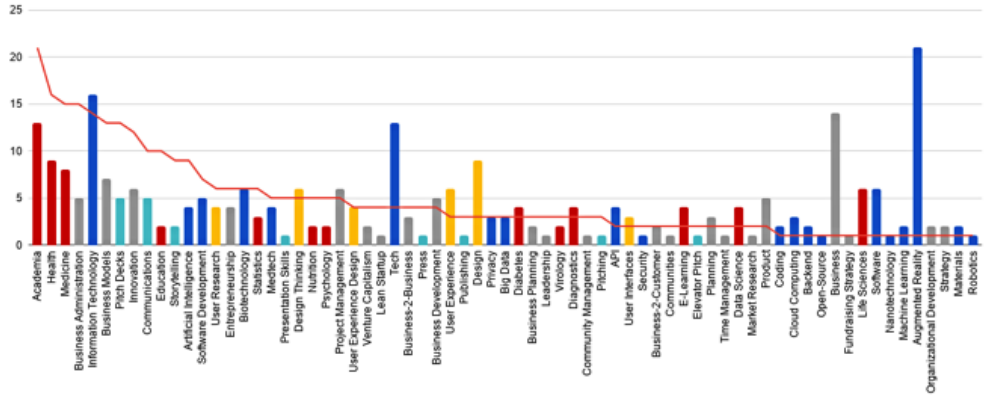
experience) design. Half of the participants (n=24) had never joined a hackathon before.

Mentoring

A total of 60 mentors preregistered online over a period of 6 days prior to the hackathon. Of them, 52 completed their profiles and became active on the Mentornity platform, of whom 46 joined Slack to communicate with teams and organizers. A briefing session was held to inform mentors about the online hackathon concept and how to offer expertise and express their availability to participants. In total, 75 video mentoring sessions took place on Mentornity where 35 mentors and 19 teams interacted. The actual number of video sessions is estimated to have been higher since it was known that other video conference platforms such as Zoom, Slack, and Skype were used and not tracked by Mentornity.

Of all mentors, 24 had no previous experience mentoring, 12 had experience from a previous Hacking Health Berlin hackathon, and the remaining 16 had experience mentoring but were new to Hacking Health Berlin events. The group of mentors had broad professional backgrounds as shown in Figure 2. The top fields of expertise comprised business, medicine, and software development. Along those lines, the most booked mentoring sessions reflected a high interest in these fields.

Figure 2. Sessions booked in the different fields of expertise available for mentorship (red line) and the number of mentors with knowledge in the respective disciplines (bars). Participants booked mentoring sessions in the field of tech (blue bars); health care, research, and education (red bars); entrepreneurship and management (gray bars); public speaking and public relations (teal bars); user research and design (yellow bars). API: application programming interface.



Feedback forms (Multimedia Appendix 1) were automatically created on Mentornity for each mentoring session conducted. Of the mentors, 13 completed the feedback forms, covering a total of 25 mentoring sessions involving 13 teams. The feedback showed that mentors were largely satisfied with the online mentoring experience despite the new format. Of 25 mentoring sessions, 22 took place and were evaluated with ratings ranging from 3 to 5, scoring an average of 4.68 (SD 0.57). Three surveys indicated that the mentoring session did not occur for various reasons: cancellation of the meeting, difficulties with the platform, or the teams did not attend the scheduled mentoring session. The teams that received mentoring reported gaining valuable insights, although they also highlighted that online mentoring was a new format and brought interaction and scheduling challenges. A number of positive comments by mentored teams reflected appreciation for mentors' willingness to share their experience, which enabled them to take a broader perspective of their project. However, during the hackathon, some teams and mentors reported difficulties in scheduling sessions or connecting via the platform, resulting in some frustration.

Projects

A total of 14 projects were developed and presented to the jury and public audience. The scope of the created solutions included telemedicine services, personal risk assessment tools, trackers for COVID-19–specific symptoms and mental health impairments, tracers for population density and risk contacts, immunity passports, hardware for safe reuse of personal protective equipment, platforms for sharing food and other essential products, and educational tools.

The *most innovative* solution [57] consisted of a mobile app for knowledge transfer between researchers, the public, and health care workers during and after the pandemic. It tackled the need to access clear and timely COVID-19 scientific evidence for researchers and practitioners to have a clear picture of the situation. App users were able to share information through short videos and visual graphics, aided by curated information and deep learning. The team presented a functional website presenting an novel artificial intelligence–based tool where users could enter specifics about their professional background and their individual research question in a free-text field to obtain summarized research findings from peer-reviewed scientific literature.

The *most impactful* solution [58] was a service platform with a video-based hotline to provide instant psychological support for health care professionals exposed to stress, a high workload, and moral dilemmas during the pandemic. With restricted time, health care workers lack the capacity to reach out to existing psychosocial support structures. By bringing psychological support to health care professionals directly via a video chat hotline, the instant psychological support was made possible by the voluntary commitment of a broad network of psychotherapists. With a well-planned concept presented on a functional website, the project straightforwardly addressed an imminent user need as identified by frontline health care professionals and outlined in challenge 5.

The *most applicable* solution [59] consisted of an app that allowed frontline health care workers involved in COVID-19 patient care to perform regular virtual health self-checks and seek support to ensure their mental, as well as physical health, was prioritized. The presented mock-up of the app had a simple user interface, self-checks required very little time, and support tools were easy to access.

The app rated as *most feasible* [60] helped to redistribute critical resources and services. As the pandemic has disrupted supply chains around the world, at the time plenty of resources available were not arriving at locations where needed. The platform solved the issue by allowing people to easily know what was needed, indicate where to send these resources, and track the resource shipping progress. A web portal ready for immediate implementation and use, as well as an outline of short and long-term goals of the project, was presented.

The *best executed* solution [61] focused on privacy-first contact tracing and digital epidemiology. It created a multiservice platform connecting patients to self-assessment tools, dashboards, and personal analytics. Users would be able to easily track their logins on their mobile phones and see how symptoms changed. The secure and anonymous contact tracing would allow patients to be aware in their residential area and avoid potential high-risk areas. Finally, information like location, age, and test results would be collected and anonymized into a large database for further analyses and data modeling. The team presented a working web app prototype with 3 features (self-assessment for tracking COVID-19–specific symptoms, contact tracing, warning function for population-dense areas) with a user-friendly and professional-looking user interface, as well as a clear roadmap for the future implementation of their product.

Approaches by other teams included sterilization methods to reuse protective equipment for health care workers, virtual reality visualizations of epidemiological data, and tools to support social distancing in public spaces. The results presented varied from concepts and click dummies to working prototypes of mobile apps.

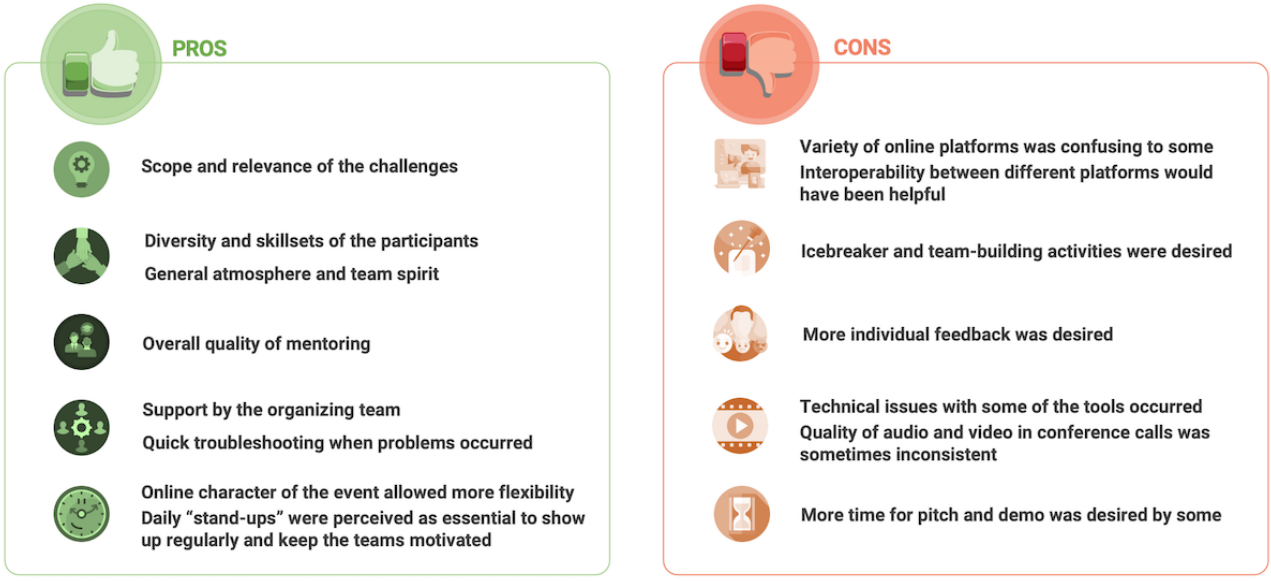
Hackathon Experience

According to the feedback (n=30), participants were made aware of the event by word of mouth (n=14, 47%), social media (n=12, 40%), event platforms (n=3, 10%), email, and postings on Hacking Health Berlin's and Data Natives' existing online communication channels on Slack (n=4, 13%).

Participants were overall satisfied with the execution of the remote hackathon, evaluating the experience with 4.23 points out of 5 on average. They appreciated the general atmosphere and team spirit, diversity and skillsets of the participants and mentors, the flexibility permitted by the remote nature of the event, and felt that occurring problems could be quickly resolved by the organizing team. However, the variety of online platforms that were used was perceived to be overwhelming to some, and technical issues with the tools were reported. Furthermore, icebreaker and team-building activities at the beginning of the event were desired to get familiar with the general concept of online hackathons and to get to know the other participants (Figure 3). Most participants (n=21, 70%) felt that it was an advantage that the hackathon was held over a long holiday weekend, but opinions were mixed about whether an in-person event of a longer duration than the usual 48 hours would be desirable, with 40% (n=12) in favor of the idea.

Before the hackathon, participants rated their confidence in starting a digital health venture or research project with 2.6 points out of 5 on average. After the hackathon, the participants' confidence increased significantly to 3.7 on average. Of the participants, 60% (n=18) would not have started their projects without attending this hackathon, and agreed that it enabled them to progress faster by (1) offering new insights and feedback provided by the mentors, (2) creating a functional prototype, and (3) finding the right members for their team.

Figure 3. Experiences of participants specifically regarding the remote and online nature of the hackathon.



Discussion

Main Findings

This case study demonstrates the feasibility and effectiveness of remote hackathons as an approach to combat the challenges of the COVID-19 pandemic. The event was prepared in a much shorter time and conducted entirely online at significantly lower costs, with similar results as health hackathons held in person. To our knowledge, this is one of the first studies to report and critically discuss experiences of, and provide guidance on, online hackathons and hackathons specifically targeted to solve the challenges around COVID-19, as well as to report on the impact of expert mentoring during health hackathons.

Cultural Change and Opportunities of Online Collaborations

The circumstances of the COVID-19 pandemic and the necessity for social distancing have led to a cultural change in work settings, communication, and collaboration. As a consequence, professional meetings and conferences are more likely to take place in an online setting. Based on the high number of participants and mentors, the medical and scientific way in which the challenges were addressed, and the feedback from participants who had experience in face-to-face hackathons, we can conclude that remote online hackathons are feasible and resemble some of the dynamics and synergies of the in-person ones. Online hackathons effectively bring people together at a significantly lower cost in comparison to a conventional in-person hackathon. The EasterHack event enabled participants not only to collaborate in an interdisciplinary manner but also across diverse geographic locations, avoiding the possible inconveniences and costs of travel. The transition toward an online and federated event allowed for a more diverse and global audience. There is substantial research showing a relation between increased diversity in teamwork leads and benefits, such as increased creativity and better problem-solving abilities [50,51].

Since the first online hackathons to fight COVID-19 were launched in March 2020, more than 50 online events summing up to 28,000 participants [62-67] have been held. The impact is evident in terms of a higher participation level and number of collaborating partners, which brings a multibeneficial bottom line of fewer logistical, financial, geographical, and time constraints, and reduced carbon dioxide output.

The Impact of Expert Mentoring

In recent years, mentoring has been shown to have a positive impact on the way participants gain insights, interact, execute multidisciplinary tasks, and build their prototypes during hackathons [3,6,8,28,36]. They act as guiding lights during the process by encouraging exploration and providing information in areas where participants have questions and interests. The main benefit is reflected in the holistic and complementary nature of each hack, often embracing different disciplines and resulting in outstanding solutions and pitches with a strong technical background (software or hardware); addressing health-related challenges pragmatically; and putting the whole idea into a well-polished business plan. On those grounds, the recruitment of expert mentors takes place depending on the scope of the hackathon.

In general, challenges tackling the COVID-19 pandemic are complex and require critical advice from biomedical experts and frontline health care professionals. These hurdles can be tackled through expert mentorship, complemented with knowledge from specialists for additional areas of expertise aiming to increase the feasibility and applicability of solutions. Availability of mentors is important; for example, health care staff and other essential workers may have commitments directly related to the pandemic, which limits their availability. With online mentoring programs, mentors can help participants with flexible and short mentorship sessions without altering their working schedules and personal time. Likewise, mentors located in different regions and time zones were able to support participants, which in an in-person format would not have been the case. The interdisciplinary nature and outcome of the event was reflected in the diversity of sessions booked; this also confirms the relevance of mentoring and our endeavor to have medical expertise as the backbone. For these reasons, we strongly recommend online mentorship to be considered in both virtual and on-site hackathons.

Challenges of Conducting Hackathons Remotely

Online hackathons might have some limitations such as data security concerns when it comes to personal information, especially with video conference tools, communication breakdowns, technical issues, and access to essential hardware including microphones and cameras. Only recently, the video conference service Zoom reported security issues concerning unwanted guests who disturbed some virtual meeting rooms [67-69]. While the preparation time for online hackathons can be relatively short, all tools and technical solutions need to be tested thoroughly in advance to avoid frustration among organizers, participants, and mentors. Technical glitches may also affect the ability to communicate properly, share announcements, recruit survey participants, and marketing. Furthermore, participants who work remotely might become distracted more easily with the screen being the only visual feedback platform and other work-related and personal commitments at the same time. Therefore, a certain level of media competence is recommended. Online team building activities require different approaches compared to events in physical presence. In addition, the dynamics of networking and small talk might be different. Although participants of EasterHack reported positive experiences regarding group synergies and team spirit during the event, there were 14 individuals who did not complete their projects. Further research should address drop-out reasons and challenges of attending online hackathons from the participant perspective, and elaborate on how these challenges might be resolved.

Limitations

This case study was based on data that were captured in a single online event and not directly compared with in-person events. Although participants were from various geographical locations, the sample size of 48 participants limits broad generalizations. Further studies in the course of online hackathons should be conducted to validate the stated findings.

Conclusions

Online hackathons offer an ideal setting for rapid and goal-oriented collaborative work, and a format to react quickly to urgent and unknown situations like the COVID-19 pandemic.

EasterHack participants generally reported a positive experience during the online hackathon format, increased confidence, and a higher likelihood of continuing to work on their projects. Despite remotely held events posing challenges to organizers, participants, and mentors, they can be executed much faster and at lower costs compared to in-person meetings, as well as foster team diversity and increase cross-regional collaboration. Setting up a research project and creating IT solutions within health care or research facilities can be a time-consuming and onerous process; however, concept-to-solution processes can be achieved during a single online hackathon weekend. As highlighted by an avid supporter of hackathons [70]:

> *Governments need to change. Institutions need to change. But, in the meantime, we have hackathons, where you can rapidly prototype. [...] Do you get a perfect solution? No. But anything is better than nothing. Done is better than perfect.*

Acknowledgments

Our thanks and appreciation go out to the volunteers of Hacking Health Berlin, Data Natives, and Charité – Universitätsmedizin Berlin, as well as to our mentors and judges, who have greatly contributed to this event. The hackathon received funding from the Berlin Institute of Health, Diabetes Centre Berne, and Paul Hartmann. We further acknowledge support from the German Research Foundation and the Open Access Publication Funds of Charité – Universitätsmedizin Berlin for the publication of this paper.

Authors KB and A-SP are fellows of the Charité Digital Clinician Scientist Program funded by the Charité – Universitätsmedizin Berlin and the Berlin Institute of Health. This program was initiated and led by Prof. Dr. Duska Dragun to enable resident physicians to pursue a career in digital health and academic research. With great sadness we have received the news that Prof. Dragun passed away on December 28th of 2020. This publication is dedicated to her as a mentor, friend, role model and stellar scientist.

Authors' Contributions

KB, ASP, and JH performed the literature search. KB, AL, and PDR designed the surveys and collected the data. KB analyzed the data. KB, AVS, JH, ASP, PDR, and AL wrote the initial draft of the paper, supported by ST, FB, SL, and VK. All authors have critically revised the paper and approved its final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Postsession feedback form for mentors to respond to on the Mentornity platform.
[DOCX File , 13 KB - jmir_v23i2e25283_app1.docx]

Multimedia Appendix 2
Feedback survey for hackathon participants.
[DOCX File , 15 KB - jmir_v23i2e25283_app2.docx]

Multimedia Appendix 3
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 19 KB - jmir_v23i2e25283_app3.docx]

References

1. Impact of COVID-19 on people's livelihoods, their health and our food systems: Joint statement by ILO, FAO, IFAD and WHO. World Health Organization. 2020 Oct 13. URL: https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people's-livelihoods-their-health-and-our-food-systems#:~:text=The%20economic%20and%20social%20disruption,the%20end%20of%20the%20year [accessed 2020-10-18]
2. Cardwell FS, Bisung E, Clarke AE, Elliott SJ. Hacking systemic lupus erythematosus (SLE): outcomes of the Waterlupus hackathon. Health Promot Chronic Dis Prev Can 2020 Aug;40(7-8):235-244 [FREE Full text] [doi: 10.24095/hpcdp.40.7/8.03] [Medline: 32667880]
3. Ruzgar NM, Ahuja C, Yu KE, Sallam A, Rosenthal R, Killelea B. How We Do It: Creation of a Workforce Development-Focused Track at a Surgical Hackathon. J Surg Educ 2020 Sep;77(5):1028-1032. [doi: 10.1016/j.jsurg.2020.04.005] [Medline: 32409286]

4. Ahmed AE, Mpangase PT, Panji S, Baichoo S, Souilmi Y, Fadlelmola FM, et al. Organizing and running bioinformatics hackathons within Africa: The H3ABioNet cloud computing experience. AAS Open Res 2018 Apr 18;1:9. [doi: 10.12688/aasopenres.12847.1]
5. Schoeb D, Suarez-Ibarrola R, Hein S, Dressler FF, Adams F, Schlager D, et al. Use of Artificial Intelligence for Medical Literature Search: Randomized Controlled Trial Using the Hackathon Format. Interact J Med Res 2020 Mar 30;9(1):e16606 [FREE Full text] [doi: 10.2196/16606] [Medline: 32224481]
6. Poncette A, Rojas P, Hofferbert J, Valera Sosa A, Balzer F, Braune K. Hackathons as Stepping Stones in Health Care Innovation: Case Study With Systematic Recommendations. J Med Internet Res 2020 Mar 24;22(3):e17004 [FREE Full text] [doi: 10.2196/17004] [Medline: 32207691]
7. Mitra S, Ashby J, Muhumuza A, Ndayishimiye I, Wasserman I, Santhirapala V, et al. Surgathon: a new model for creating a surgical innovation ecosystem in low-resource settings. BMJ Glob Health 2020 Feb 16;5(2):e002162 [FREE Full text] [doi: 10.1136/bmjgh-2019-002162] [Medline: 32133197]
8. Li C, Xiong Y, Sit HF, Tang W, Hall BJ, Muessig KE, et al. A Men Who Have Sex With Men-Friendly Doctor Finder Hackathon in Guangzhou, China: Development of a Mobile Health Intervention to Enhance Health Care Utilization. JMIR Mhealth Uhealth 2020 Feb 27;8(2):e16030 [FREE Full text] [doi: 10.2196/16030] [Medline: 32130189]
9. Pathanasethpong A, Areemit R, Teerakulpisut D, Morley K, Morley M. Health hackathon as a venue for interprofessional education: a qualitative interview study. J Interprof Care 2019 Dec 21;34(6):1-3. [doi: 10.1080/13561820.2019.1696760] [Medline: 31865825]
10. Ferreira GC, Oberstaller J, Fonseca R, Keller TE, Adapa SR, Gibbons J, et al. Iron Hack - A symposium/hackathon focused on porphyrias, Friedreich's ataxia, and other rare iron-related diseases. F1000Res 2019 Jul 19;8:1135 [FREE Full text] [doi: 10.12688/f1000research.19140.1] [Medline: 31824661]
11. McLeod P, Cunningham Q, DiazGranados D, Dodoiu G, Kaplan S, Keyton J, et al. Hacking teamwork in health care: Addressing adverse effects of ad hoc team composition in critical care medicine. Health Care Manage Rev 2019 Dec 03:A. [doi: 10.1097/HMR.0000000000000265] [Medline: 31804232]
12. Fecho K, Ahalt SC, Arunachalam S, Champion J, Chute CG, Davis S, Biomedical Data Translator Consortium. Sex, obesity, diabetes, and exposure to particulate matter among patients with severe asthma: Scientific insights from a comparative analysis of open clinical data sources during a five-day hackathon. J Biomed Inform 2019 Dec;100:103325 [FREE Full text] [doi: 10.1016/j.jbi.2019.103325] [Medline: 31676459]
13. Connor R, Brister R, Buchmann J, Deboutte W, Edwards R, Martí-Carreras J, et al. NCBI's Virus Discovery Hackathon: Engaging Research Communities to Identify Cloud Infrastructure Requirements. Genes (Basel) 2019 Sep 16;10(9):714 [FREE Full text] [doi: 10.3390/genes10090714] [Medline: 31527408]
14. Panchapakesan C, Sheldenkar A, Wimalaratne P, Wijayamuni R, Lwin MO. Developing a Digital Solution for Dengue Through Epihack: Qualitative Evaluation Study of a Five-Day Health Hackathon in Sri Lanka. JMIR Form Res 2019 Aug 29;3(3):e11555 [FREE Full text] [doi: 10.2196/11555] [Medline: 31469074]
15. Ramadi KB, Srinivasan S, Atun R. Health diplomacy through health entrepreneurship: using hackathons to address Palestinian-Israeli health concerns. BMJ Glob Health 2019 Jul 10;4(4):e001548. [doi: 10.1136/bmjgh-2019-001548]
16. Bell J, Murray F, Davies E. An investigation of the features facilitating effective collaboration between public health experts and data scientists at a hackathon. Public Health 2019 Aug;173:120-125. [doi: 10.1016/j.puhe.2019.05.007] [Medline: 31271966]
17. Lwin MO, Sheldenkar A, Panchapakesan C, Ng JS, Lau J, Jayasundar K, et al. Epihack Sri Lanka: development of a mobile surveillance tool for dengue fever. BMC Med Inform Decis Mak 2019 Jun 13;19(1):111 [FREE Full text] [doi: 10.1186/s12911-019-0829-5] [Medline: 31196073]
18. Aungst TD, Patel R, Pugliese R, Patel I, Boutari C. From ideation to practice: How pharmacists and students can leverage hackathons and innovation labs to accelerate innovation in pharmacy. J Am Pharm Assoc (2003) 2019 Mar;59(2S):S25-S29. [doi: 10.1016/j.japh.2018.12.009] [Medline: 30713082]
19. Wang JK, Pamnani RD, Capasso R, Chang RT. Correction to: An Extended Hackathon Model for Collaborative Education in Medical Innovation. J Med Syst 2018 Dec 10;43(1):15. [doi: 10.1007/s10916-018-1127-y] [Medline: 30536040]
20. Wang JK, Roy SK, Barry M, Chang RT, Bhatt AS. Institutionalizing healthcare hackathons to promote diversity in collaboration in medicine. BMC Med Educ 2018 Nov 20;18(1):269 [FREE Full text] [doi: 10.1186/s12909-018-1385-x] [Medline: 30458759]
21. Tucci C, Viscusi G, Gautschi H. Translating Science Into Business Innovation: The Case of Open Food and Nutrition Data Hackathons. Front Nutr 2018 Oct 22;5:96 [FREE Full text] [doi: 10.3389/fnut.2018.00096] [Medline: 30406107]
22. Wang JK, Pamnani RD, Capasso R, Chang RT. An Extended Hackathon Model for Collaborative Education in Medical Innovation. J Med Syst 2018 Oct 17;42(12):239. [doi: 10.1007/s10916-018-1098-z] [Medline: 30328518]
23. Brown A, Jauregui J, Chipman A, Ilgen J. Using Hackathons to Transform Complex Educational Problems Into Innovative Prototypes. J Grad Med Educ 2018 Aug;10(4):465-466 [FREE Full text] [doi: 10.4300/JGME-D-18-00395.1] [Medline: 30154981]

24. Ghouila A, Siwo GH, Entfellner JD, Panji S, Button-Simons KA, Davis SZ, DREAM of Malaria Hackathon Participants, et al. Hackathons as a means of accelerating scientific discoveries and knowledge transfer. Genome Res 2018 May 12;28(5):759-765 [FREE Full text] [doi: 10.1101/gr.228460.117] [Medline: 29650552]
25. Lyndon MP, Cassidy MP, Celi LA, Hendrik L, Kim YJ, Gomez N, et al. Hacking Hackathons: Preparing the next generation for the multidisciplinary world of healthcare technology. Int J Med Inform 2018 Apr;112:1-5. [doi: 10.1016/j.ijmedinf.2017.12.020] [Medline: 29500006]
26. Cooper K, Siefert A, Weinreb J. Skills Beyond the Reading Room: Training in Innovation and Collaboration at a Radiology Hackathon. J Am Coll Radiol 2018 Mar;15(3 Pt A):466-468. [doi: 10.1016/j.jacr.2017.10.025] [Medline: 29305074]
27. Firenze A, Provenzano S, Santangelo O, Alagna E, Piazza D, Torregrossa M. Hackathon Public Health. Clin Ter 2017;168(6):e421-e427 [FREE Full text] [doi: 10.7417/T.2017.2045] [Medline: 29209692]
28. Pathanasethpong A, Soomlek C, Morley K, Morley M, Polpinit P, Dagan A, et al. Tackling Regional Public Health Issues Using Mobile Health Technology: Event Report of an mHealth Hackathon in Thailand. JMIR Mhealth Uhealth 2017 Oct 16;5(10):e155 [FREE Full text] [doi: 10.2196/mhealth.8259] [Medline: 29038098]
29. Madhani S, Farooqi WH, Mian AI. Stimulating innovation through the hackathon concept in paediatrics: our experience at the Aga Khan University. Arch Dis Child 2017 Oct 16;102(10):994-994. [doi: 10.1136/archdischild-2017-313648] [Medline: 28814421]
30. Kohli M, Morrison JJ, Wawira J, Morgan MB, Hostetter J, Genereaux B, et al. Creation and Curation of the Society of Imaging Informatics in Medicine Hackathon Dataset. J Digit Imaging 2018 Feb 20;31(1):9-12 [FREE Full text] [doi: 10.1007/s10278-017-0003-5] [Medline: 28730549]
31. hackseq Organizing Committee 2016. Catalyzing collaboration between biological and computational scientists via hackathon. F1000Res 2017 Feb 28;6:197 [FREE Full text] [doi: 10.12688/f1000research.10964.2] [Medline: 28417000]
32. Angelidis P, Berman L, Casas-Perez MDLL, Celi LA, Dafoulas GE, Dagan A, et al. The hackathon model to spur innovation around global mHealth. J Med Eng Technol 2016 Aug 10;40(7-8):392-399 [FREE Full text] [doi: 10.1080/03091902.2016.1213903] [Medline: 27538360]
33. Kapur T, Pieper S, Fedorov A, Fillion-Robin J, Halle M, O'Donnell L, et al. Increasing the impact of medical image computing using community-based open-access hackathons: The NA-MIC and 3D Slicer experience. Med Image Anal 2016 Oct;33:176-180 [FREE Full text] [doi: 10.1016/j.media.2016.06.035] [Medline: 27498015]
34. Silver JK, Binder DS, Zubcevik N, Zafonte RD. Healthcare Hackathons Provide Educational and Innovation Opportunities: A Case Study and Best Practice Recommendations. J Med Syst 2016 Jul 8;40(7):177 [FREE Full text] [doi: 10.1007/s10916-016-0532-3] [Medline: 27277278]
35. Verschoore JR. coordenação de esforços coletivos para enfrentar a pandemia do novo coronavírus: um estudo de caso sobre o hackathon hack for Brazil | COVID-19. REAd Rev eletrôn adm (Porto Alegre) 2020 Aug;26(2):238-264. [doi: 10.1590/1413-2311.283.103388]
36. Bacq S, Geoghegan W, Josefy M, Stevenson R, Williams TA. The COVID-19 Virtual Idea Blitz: Marshaling social entrepreneurship to rapidly respond to urgent grand challenges. Bus Horiz 2020 Nov;63(6):705-723 [FREE Full text] [doi: 10.1016/j.bushor.2020.05.002] [Medline: 32398883]
37. Song Y, Zhang M, Yin L, Wang K, Zhou Y, Zhou M, et al. COVID-19 treatment: close to a cure? A rapid review of pharmacotherapies for the novel coronavirus (SARS-CoV-2). Int J Antimicrob Agents 2020 Aug;56(2):106080 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.106080] [Medline: 32634603]
38. Esakandari H, Nabi-Afjadi M, Fakkari-Afjadi J, Farahmandian N, Miresmaeili S, Bahreini E. A comprehensive review of COVID-19 characteristics. Biol Proced Online 2020 Aug 4;22(1):19 [FREE Full text] [doi: 10.1186/s12575-020-00128-2] [Medline: 32774178]
39. Tang Y, Liu J, Zhang D, Xu Z, Ji J, Wen C. Cytokine Storm in COVID-19: The Current Evidence and Treatment Strategies. Front Immunol 2020 Jul 10;11:1708 [FREE Full text] [doi: 10.3389/fimmu.2020.01708] [Medline: 32754163]
40. Kai H, Kai M. Interactions of coronaviruses with ACE2, angiotensin II, and RAS inhibitors-lessons from available evidence and insights into COVID-19. Hypertens Res 2020 Jul 27;43(7):648-654 [FREE Full text] [doi: 10.1038/s41440-020-0455-8] [Medline: 32341442]
41. Sanders J, Monogue M, Jodlowski T, Cutrell J. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. JAMA 2020 May 12;323(18):1824-1836. [doi: 10.1001/jama.2020.6019] [Medline: 32282022]
42. Pascarella G, Strumia A, Piliego C, Bruno F, Del Buono R, Costa F, et al. COVID-19 diagnosis and management: a comprehensive review. J Intern Med 2020 Aug 13;288(2):192-206 [FREE Full text] [doi: 10.1111/joim.13091] [Medline: 32348588]
43. Al-Quteimat O, Amer A. The Impact of the COVID-19 Pandemic on Cancer Patients. Am J Clin Oncol 2020 Jun;43(6):452-455 [FREE Full text] [doi: 10.1097/COC.0000000000000712] [Medline: 32304435]
44. Zhou M, Xie X, Peng Y, Wu M, Deng X, Wu Y, et al. From SARS to COVID-19: What we have learned about children infected with COVID-19. Int J Infect Dis 2020 Jul;96:710-714 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.090] [Medline: 32389849]

45. Godri Pollitt KJ, Peccia J, Ko AI, Kaminski N, Dela Cruz CS, Nebert DW, et al. COVID-19 vulnerability: the potential impact of genetic susceptibility and airborne transmission. Hum Genomics 2020 May 12;14(1):17 [FREE Full text] [doi: 10.1186/s40246-020-00267-3] [Medline: 32398162]
46. Randolph HE, Barreiro LB. Herd Immunity: Understanding COVID-19. Immunity 2020 May 19;52(5):737-741 [FREE Full text] [doi: 10.1016/j.immuni.2020.04.012] [Medline: 32433946]
47. Ali I, Alharbi OM. COVID-19: Disease, management, treatment, and social impact. Sci Total Environ 2020 Aug 01;728:138861 [FREE Full text] [doi: 10.1016/j.scitotenv.2020.138861] [Medline: 32344226]
48. Sood S, Aggarwal V, Aggarwal D, Upadhyay SK, Sak K, Tuli HS, et al. COVID-19 Pandemic: from Molecular Biology, Pathogenesis, Detection, and Treatment to Global Societal Impact. Curr Pharmacol Rep 2020 Jul 27;6(5):1-16 [FREE Full text] [doi: 10.1007/s40495-020-00229-2] [Medline: 32837855]
49. Hacking Health - EasterHack. 2020. URL: https://hacking-health.org/easter-hackathon2020/ [accessed 2020-10-26]
50. Usher M, Barak M. Team diversity as a predictor of innovation in team projects of face-to-face and online learners. Computers & Education 2020 Jan;144:103702. [doi: 10.1016/j.compedu.2019.103702]
51. Bouncken R, Brem A, Kraus S. Multi-cultural teams as sources for creativity and innovation: the role of cultural diversity on team performance. Int J Innov Mgt 2016 Feb 15;20(01):1650012. [doi: 10.1142/s1363919616500122]
52. LinkedIn. URL: http://www.linkedin.com [accessed 2020-11-20]
53. Slack. URL: http://www.slack.com/ [accessed 2020-11-18]
54. Airtable. URL: http://www.airtable.com/ [accessed 2020-11-18]
55. Mentornity. URL: http://www.mentornity.com/ [accessed 2020-11-18]
56. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
57. THM - The Health Manager. URL: https://hackcorona-berlin.sparkboard.com/project/5e9033478757d6002407f0d7 [accessed 2020-12-15]
58. Psychische Unterstützung während Krisenzeiten. mAIndcraft. URL: https://hotline.maindcraft.com/ [accessed 2020-12-15]
59. MyCare. URL: https://hackcorona-berlin.sparkboard.com/project/5e90a7424a73ae002467088a [accessed 2020-12-15]
60. contribute2.us. URL: https://hackcorona-berlin.sparkboard.com/project/5e90a9d74a73ae002467088f [accessed 2020-12-15]
61. reCOVer. URL: https://hackcorona-berlin.sparkboard.com/project/5e9001868757d6002407f0cf [accessed 2020-12-15]
62. Pan-European Matchathon. URL: http://www.euvsvirus.org [accessed 2020-10-18]
63. #WirVsVirus - #WirVSVirus. URL: http://www.wirvsvirus.org [accessed 2020-10-18]
64. Talbert T. Space Apps COVID-19 Hackathon Brings 15,000 Together Worldwide. NASA. 2020 Jun 9. URL: https://www.nasa.gov/feature/space-apps-covid-19-hackathon-brings-15000-together-worldwide [accessed 2020-10-18]
65. Kinder F, Harvey A. Covid-19: the medical students responding to the pandemic. BMJ 2020 Jun 15;369:m2160. [doi: 10.1136/bmj.m2160] [Medline: 32540952]
66. Luo EM, Newman S, Amat M, Charpignon M, Duralde ER, Jain S, et al. MIT COVID-19 Datathon: data without boundaries. BMJ Innov 2021 Jan 31;7(1):231-234 [FREE Full text] [doi: 10.1136/bmjinnov-2020-000492] [Medline: 33437494]
67. Wagenseil P. Zoom security issues: Here's everything that's gone wrong (so far). Tom's Guide. 2021 Jan 5. URL: https://www.tomsguide.com/news/zoom-security-privacy-woes [accessed 2021-01-28]
68. Paul K. 'Zoom is malware': why experts worry about the video conferencing platform. The Guardian. 2020 Apr 2. URL: http://www.theguardian.com/technology/2020/apr/02/zoom-technology-security-coronavirus-video-conferencing [accessed 2020-10-18]
69. O'Flaherty K. Zoom Security: 100 New Features Later, Can You Trust Zoom? Forbes. 2020 Jul 2. URL: https://www.forbes.com/sites/kateoflahertyuk/2020/07/02/zoom-security-100-new-features-later-can-you-trust-zoom-yet/ [accessed 2020-10-18]
70. Garfinkel J. Hacking COVID-19. The Walrus. 2020 Oct 7. URL: https://thewalrus.ca/hacking-covid-19 [accessed 2020-10-18]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
IT: information technology
UI: user interface
UX: user experience

Edited by G Eysenbach; submitted 23.11.20; peer-reviewed by H Park, A Al-Hindawi; comments to author 14.12.20; revised version received 02.01.21; accepted 16.01.21; published 08.02.21.

Please cite as:
Braune K, Rojas PD, Hofferbert J, Valera Sosa A, Lebedev A, Balzer F, Thun S, Lieber S, Kirchberger V, Poncette AS
Interdisciplinary Online Hackathons as an Approach to Combat the COVID-19 Pandemic: Case Study
J Med Internet Res 2021;23(2):e25283
URL: http://www.jmir.org/2021/2/e25283/
doi:10.2196/25283
PMID:33497350

Original Paper

Early Perceptions of COVID-19 Contact Tracing Apps in German-Speaking Countries: Comparative Mixed Methods Study

Bettina Maria Zimmermann[1,2], Dr phil des; Amelia Fiske[1], PhD; Barbara Prainsack[3], Prof Dr; Nora Hangel[1], PhD; Stuart McLennan[1,2*], PhD; Alena Buyx[1*], MD, PhD

[1]Institute of History and Ethics in Medicine, Technical University Munich, Munich, Germany

[2]Institute for Biomedical Ethics, University of Basel, Basel, Switzerland

[3]Department of Political Science, University of Vienna, Vienna, Austria

[*]these authors contributed equally

Corresponding Author:
Bettina Maria Zimmermann, Dr phil des
Institute of History and Ethics in Medicine
Technical University Munich
Ismaninger Straße 22
Munich,
Germany
Phone: 49 89 4140 4042
Email: bettina.zimmermann@tum.de

Abstract

Background: The main German-speaking countries (Germany, Austria, and Switzerland) have implemented digital contact tracing apps to assist the authorities with COVID-19 containment strategies. Low user rates for these apps can affect contact tracing and, thus, its usefulness in controlling the spread of the novel coronavirus.

Objective: This study aimed to assess the early perceptions of people living in the German-speaking countries and compare them with the frames portrayed in the newspapers during the first wave of the COVID-19 pandemic.

Methods: We conducted qualitative interviews with 159 participants of the SolPan project. Of those, 110 participants discussed contact tracing apps and were included in this study. We analyzed articles regarding contact tracing apps from 12 newspapers in the German-speaking countries.

Results: Study participants perceived and newspaper coverage in all German-speaking countries framed contact tracing apps as governmental surveillance tools and embedded them in a broader context of technological surveillance. Participants identified trust in authorities, respect of individual privacy, voluntariness, and temporary use of contact tracing apps as prerequisites for democratic compatibility. Newspapers commonly referenced the use of such apps in Asian countries, emphasizing the differences in privacy regulation among these countries.

Conclusions: The uptake of digital contact tracing apps in German-speaking countries may be undermined due to privacy risks that are not compensated by potential benefits and are rooted in a deeper skepticism towards digital tools. When authorities plan to implement new digital tools and practices in the future, they should be very transparent and proactive in communicating their objectives and the role of the technology—and how it differs from other, possibly similar, tools. It is also important to publicly address ethical, legal, and social issues related to such technologies prior to their launch.

(J Med Internet Res 2021;23(2):e25525) doi:10.2196/25525

KEYWORDS

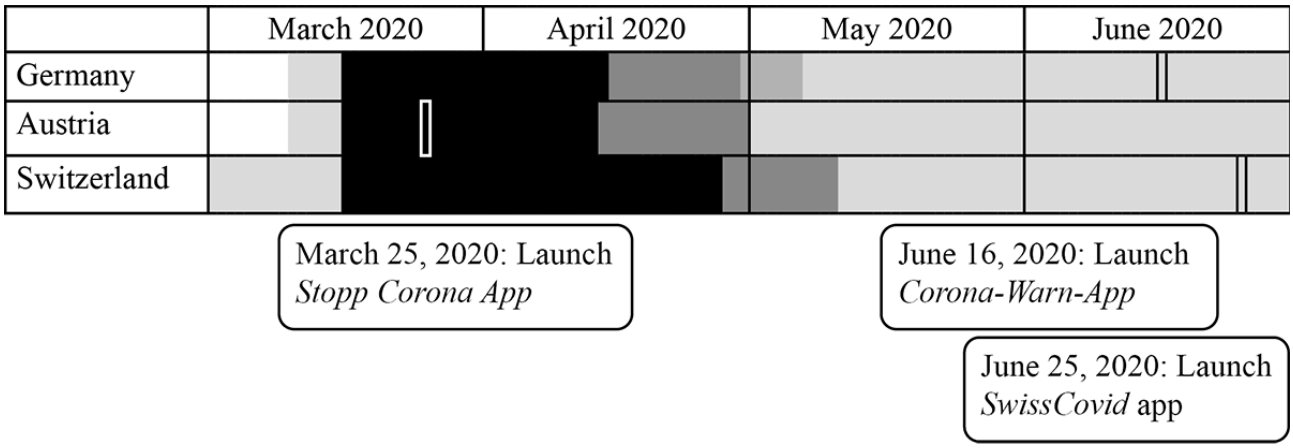
COVID-19; contact tracing; app; interview study; newspaper content analysis; privacy paradox; digital surveillance; trust; content analysis; surveillance; privacy; interview

Introduction

During the COVID-19 pandemic, several country-specific contact tracing apps were introduced to assist and supplement analog contact-tracing. In Austria, Red Cross launched the Stopp Corona app on March 25, 2020; Germany launched the Corona-Warn-App on June 16, 2020; and Switzerland launched the SwissCovid app on June 25, 2020 (Figure 1). All these free

apps use Bluetooth Low Energy technology to measure the distance and the duration of contact between smartphones that have the app installed. If the tracking function is enabled, random encrypted codes are automatically exchanged and directly saved on the devices whenever a person encounters another user. These random codes do not reveal any names or identities of people or exact locations, and the codes are erased from the smartphones after 14 days. If a person using the app tests positive for COVID-19, they can voluntarily choose to make their random codes anonymously available so that other users who may have come into contact with the infected person can be notified and asked to contact their local health authorities for further instructions.

Figure 1. Comparative timeline of COVID-19–related restrictions and contact tracing smartphone apps released from March to June 2020. Black: lockdown period; grey: step-wise easing of restrictions.



Although these apps have the same basic functions, there are some important differences among them. First, there are differences concerning the institution that launched the apps—the Swiss and German contact tracing apps were launched by the respective federal governments, whereas the Austrian app was launched by Red Cross Austria, a national nongovernmental organization. Second, the apps differ in the way in which users can report infection. For instance, German app users need to call a hotline after having received a positive COVID-19 test result [1]. Swiss app users receive a code from health authorities that needs to be entered into the app [2]. Austrian app users can alert their contacts directly without contacting any authority, but they need to register their phone numbers to avoid misuse [3]. The Austrian app also includes a symptom checker that allows for a direct alert if symptoms indicative of COVID-19 are entered [3]. Third, there are differences concerning the warning levels: the Swiss and Austrian apps have 2 warning levels (“no warnings” or “potential infection”), whereas the German app has an additional option called “unknown risk,” which is useful in cases wherein the app was not activated long enough to make an assessment. Finally, there are also differences in the management systems used by these apps. In Europe, key management systems for such contact tracing apps include Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT) [4] and Decentralized Privacy-Preserving Proximity Tracing (DP-3T) [5]. PEPP-PT aimed for a centralized data storage system, whereas DP-3T foresaw a decentralized data storage approach on individual phones, thereby aiming to increase data security and privacy [6,7]. The final versions of all 3 apps used the decentralized approach as promoted by DP-3T.

However, it appears that very few people have downloaded and are using these apps. Although surveys conducted in Spring 2020 reported acceptance rates of up to 70% for hypothetical contact tracing apps in Germany and Switzerland [8,9], only 25% and 31% of the total populations in Germany and Switzerland, respectively, had downloaded the app by October 30, 2020, with active user estimations ranging around 21% in both countries [2,10]. On the other hand, only 12% of the population in Austria had downloaded the Stopp Corona App by mid-April 2020 [11,12], and download rates only increased marginally to 12.4% until October 2020 [13]. Modeling studies estimated that for contact tracing apps to be effective, the user rate needed to be at least 56% of the overall population [14] and that uptake rates need to be higher for decentralized data storage systems than for centralized ones [15]. Nevertheless, even lower user rates can have an impact on contact tracing and, consequently, on the ability of the system to control the spread of COVID-19 [16]. Contrary to those estimates, Maccari and Cagno [17] argued that Bluetooth-based, privacy-preserving contact tracing apps are of limited benefit due to potentially high false-positive rates and low sensitivity. Similarly, Rowe et al [18] contended that the French contact tracing app was implemented very quickly without taking into account the available evidence on viral spread. Overall, robust evidence on the actual effectiveness of contact tracing apps is lacking [19].

A large body of the existing literature has assessed users’ perceptions, motivations, and barriers to using mobile health (mHealth) apps, particularly those related to chronic diseases and behavioral interventions [20-22]. Previous research on peoples’ motivation to use mHealth apps suggests that user’s perceived personal value is a crucial motivating factor [23-27], but other factors such as convenient use [26], high quality, trustworthy information provided through the apps [28], and previous experience and training with mobile technology particularly among older adults [27,29] are also mentioned. Stach et al [30] recently suggested a framework to evaluate mHealth apps, including user engagement, app functionality, aesthetics, information quality, therapeutic gain, users’ subjective quality ratings, and perceived impact. Although

research on mHealth apps generally focuses on individual usability, including the specific needs of patients with chronic conditions [22,31] or lifestyle support for a healthy diet or physical activity [32,33], contact tracing apps for pandemic control have an additional societal component. Accordingly, in the German context, Trang et al [34] found that appeals to the societal rather than personal benefit of contact tracing apps were most helpful to maximize uptake, particularly if the majority of the population was critical of the apps or undecided about using them. Other survey studies on the willingness to use contact tracing apps indicated that perceived benefit, expected performance, trust in government, and social influence were important motivating factors for using contact tracing apps, whereas privacy concerns were identified as the main hindering factor [8,35-40]. In line with these surveys, Keshet [41] identified a dichotomy in the comments related to Israeli news websites between supporters of contact tracing apps who see them as a protective measure for containment and opponents who see their civil rights violated. Despite issues pertaining to privacy and other civil rights being the most important concern against the use of contact tracing apps [34,42,43], the privacy policies of these apps are not comprehensible by the average individual owing to their complexity [44]. Lack of privacy policies have also been identified for other mobile health apps [45], mHealth apps [46] and privacy concerns, including data safety and confidentiality, are known to negatively affect the uptake of mHealth apps in general [20,46-48].

How digital contact tracing apps are perceived will ultimately be a key factor in people's willingness to download and use them. Therefore, there is a need to better understand people's views on the use of digital technology to support COVID-19 contact tracing. The COVID-19 pandemic has been widely covered in the media, mirroring the immense public interest, with traditional mass media, including print and online news media, considered to be one of the preferred information sources [49,50]. The dynamic and uncertain situation at the beginning of the pandemic made people particularly reliant on such information sources. Previous studies have used media content analyses to assess public debates during the COVID-19 pandemic [18,50], as well as other issues that affect population health [51,52]. Some other studies have combined media content analyses with surveys and interviews [53,54].

This study aimed to examine the early perceptions of people living in Germany, Austria, and German-speaking Switzerland toward digital contact tracing apps and compare them with frames as portrayed in the newspapers during the first wave of the COVID-19 pandemic. More specifically, the following research questions were addressed: (1) How did people living in Germany, Austria, and German-speaking Switzerland conceptualize and evaluate digital contact tracing apps during the COVID-19 lockdown in April 2020? (2) How were such apps portrayed in the newspapers, and were there any country-specific differences observed? (3) How did people's concepts and assessments intersect with the ongoing public debates and policies in April 2020? The countries investigated in this study are quite similar in terms of their culture (eg, language and a strong emphasis on privacy) and have democratic federalist political systems. By comparing data from similar countries, the observed differences can be interpreted in a more precise way [55]. In this context, one relevant difference is the circumstance that Austria had already launched a contact tracing app at the time of the interviews of this study, whereas Switzerland and Germany had not.

Methods

Qualitative Interviews

As part of the qualitative, longitudinal, and multinational SolPan (Solidarity in Times of a Pandemic) study [56], qualitative interviews were conducted with 159 individuals in Germany, Austria, and German-speaking Switzerland during the first COVID-19–related lockdown of April 2020. The SolPan Consortium includes 9 European countries (Austria, Belgium, Germany, France, Ireland, Italy, The Netherlands, German-speaking Switzerland, and the United Kingdom). It aims to explore peoples' experiences during the COVID-19 pandemic, particularly how people describe practices relating to solidarity. Interviews were conducted between April 6 and May 6, 2020, during which country-specific measures were in place to flatten the infection curve. This included various restrictions on movement and contact with other individuals, and the closure of schools, nonessential businesses, and public institutions.

Participants were recruited through online advertisement via university websites, social media networks, convenience sampling, and snowballing. To enable a variety of perspectives, participants were recruited with attention to a range of different demographics, including age, gender, income, household structure, geographic area, education, and employment. Participants received a study information leaflet prior to the interview, and verbal consent was obtained directly before the interview. A researcher-developed interview guide was used to guide the interview, which included a question regarding the use of mobile phones to assist in contact tracing (see Multimedia Appendix 1 for the interview guide). Interviews ranged from 25 minutes to 80 minutes in length; they were conducted online or by telephone and recorded on a digital recorder or using a video chat recorder compliant with the European General Data Protection Regulation. Only audio material was stored. The interviews were transcribed and subsequently pseudonymized. The study was approved by the ethics committees of the Technical University Munich (no. 208/20 S) and the University of Vienna (no. 00544). Interview transcripts were coded by all researchers by using an inductively generated coding scheme developed through the broader SolPan Consortium, using the ATLAS.ti 8.0 software (ATLAS.ti Scientific Software Development GmbH). Coding was checked by a second researcher for consistency. Relevant text passages were extracted using the Atlas.ti query function, analyzed inductively, and summarized in a memo by the first author (BZ), thereby aiming to gain a higher level of abstraction by building concepts and categories. This analysis was performed separately for each country, and 2 authors (SM for Germany and Switzerland, and BP for Austria) double-checked and supplemented the analysis. Then, the concepts and categories were compared between the 3 countries and discussed among the researchers. Interviews

were analyzed in German, but memos were written in English. Illustrative quotes were translated from German to English by a native German speaker (BZ) and double-checked by a native English speaker with good German skills (SM).

Content Analysis of Newspaper Coverage

In parallel to the interview analysis, a newspaper content analysis was conducted to assess what concepts, topics, and concerns were predominant at the time of the interviews across the countries. These insights were used to create a comparative framework analysis for country-specific interpretations and intercountry comparisons. The newspaper content analysis included articles published between March 15 and May 6, 2020, which includes the interview study period, and 3 weeks before the first interviews were conducted.

Three quality newspapers with a national readership and one tabloid from each country were included in the analyses. The selected newspapers are among those with the highest readership and were chosen based on equivalent functions in the respective national media systems [57], to allow for meaningful comparison. Newspapers were included to represent the mass media landscape of public debates for several reasons. First, even though newspapers lose readers to other channels, they still have considerable influence on coverage of other mass media, including social media for long-lasting issues [58]. Moreover, particularly large, high-quality newspapers are still considered a trustworthy source for reliable background information on health-related issues [25]. Finally, the publication format is reliably accessible. Accordingly, the Factiva database (Dow Jones Professional) [59] was used for a systematic search of articles that covered issues related to digital contact tracing. This database was selected because it covered the relevant newspapers, thus allowing for a unified search strategy, and was accessible to the research team. Relevant articles were retrieved with the following search algorithm using full-text search (in German): "(app OR technologie) AND (tracing OR tracking) AND (corona* OR covid-19)." Articles that reported on technology-based tracking or tracing in the context of COVID-19 were included in the study.

A codebook for the media content analysis was developed to collect the following variables: date of publication, medium, importance of topic (ie, contact tracing apps) in the article, country reference, article topics, app evaluation, and stakeholders cited (see Multimedia Appendix 2). The codebook was adapted from a previous study investigating newspaper coverage of medico-scientific issues [51]. Three coders collected these variables; they were trained in 2 online sessions wherein the codebook was discussed and tested in the team and refined upon discussion to make categories exclusive and intersubjectively understandable. For this aspect, 10 articles from each country were double-coded to identify the remaining ambiguities in the codebook. Since all instances of discordant coding could be easily resolved and were assignable to a source of uncertainty that was removed by clarifying the codebook, and because this analysis was used to supplement the qualitative analysis of the interviews, no formal reliability test was conducted. Descriptive statistics were calculated using Excel (Microsoft Corp). To identify key events, the distribution of articles over time as well as information from policy analyses were qualitatively linked to the collected variables. Key events were identified when several newspapers covered the same topic on the same day and/or when a topic was followed-up on several subsequent days.

Results

In German-speaking countries, both interview participants and newspaper coverage perceived and framed contact tracing apps predominantly as governmental surveillance tools. In the following sections, we first report the findings of the interviews, followed by results of the newspaper content analysis.

Participants' Perceptions of Contact Tracing Apps

In the 159 interviews conducted, more than two-thirds (110/159, 69.2%) of the participants commented on contact tracing apps and were included in this analysis (Germany n=29, Austria n=56, and Switzerland n=25). Table 1 presents the demographics of the study participants.

Table 1. Demographic distribution of study participants.

Characteristic	Value, n (%)		
	Germany (n=29)	Austria (n=56)	Switzerland (n=25)
Age (years)			
18-30	4 (14)	11 (20)	7 (28)
31-45	14 (56)	11 (24)	3 (17)
46-60	4 (36)	20 (59)	6 (40)
61-70	6 (29)	12 (25)	4 (17)
>70	1 (3)	2 (3)	5 (17)
Sex[a]			
Female	14 (48)	34 (61)	14 (56)
Male	15 (52)	22 (39)	11 (44)
Household			
Single	7 (24)	15 (27)	7 (28)
Couple	10 (34)	20 (36)	8 (32)
Living with child or children under 12 years	8 (28)	7 (13)	1 (4)
Living with child or children above 12 years	2 (7)	8 (14)	4 (16)
Other	2 (7)	6 (11)	5 (20)
Rural/urban			
Big town (eg, capital, >500,000 population)	12 (41)	30 (54)	8 (32)
Mid-sized or small town	9 (31)	15 (27)	5 (20)
Rural (eg, village)	8 (28)	11 (20)	12 (48)
Employment status			
Employed (long-term contract)	17 (59)	21 (38)	9 (36)
Self-employed	2 (7)	9 (16)	3 (12)
Employed (short-term or precarious contract)	1 (3)	4 (7)	5 (20)
Unemployed	4 (14)	4 (7)	1 (4)
Retired	4 (14)	14 (25)	6 (24)
Other	1 (3)	4 (7)	1 (4)
Education level			
<10 years	0 (0)	5 (9)	8 (32)
10-14 years (eg, high-school diploma)	9 (31)	18 (32)	3 (12)
Higher education	20 (69)	33 (59)	14 (56)
Household net income per month[b] **(before the COVID-19 pandemic)**			
≤€1400 (US $1693) or ≤CHF 4000 (US $4472)	3 (10)	7 (13)	3 (12)
€1401-3000 (US $1694-3628) or CHF 4001-7000 (US $4473-7826)	7 (24)	23 (41)	8 (32)
>€3000 (US $3628) or >CHF 7000 (US $7826)	19 (66)	26 (46)	14 (56)
Total interviews (N=110)	29 (26.3)	56 (50.9)	25 (22.7)

[a]Investigator observed.

[b]Self reported; categories were adjusted based on country-specific income levels.

Most participants stated that they had heard about the option of mobile phone–based contact tracing, but several participants seemed to be uncertain about the function and scope of contact tracing apps or lumped together different technologies. For instance, the concept of individual contact tracing was sometimes confused with population surveillance measures, such as anonymous mobile phone tracking to analyze population behavior during restrictions. These different conceptions of how

contact tracing could be employed were also a source of confusion and uncertainty about what mobile phone–based tracing would actually entail.

> *What I don't quite understand here is whether it´s about who is positive or simply controlling masses of people, such as [navigation systems] in traffic. That uses data to control traffic volume. I don't know exactly how it is supposed to work, that the people who are positive sign up themselves, so to say, or whether the database is maintained and you don't even know whether you are in it or not.* [Swiss participant 15]

A few participants also stated that they were not well informed about the topic at the time of the interview, indicating that they were not sure whether they could endorse the use of digital tracking tools due to their limited understanding or stating that they were aware of the tools but had not yet formed an opinion on their use.

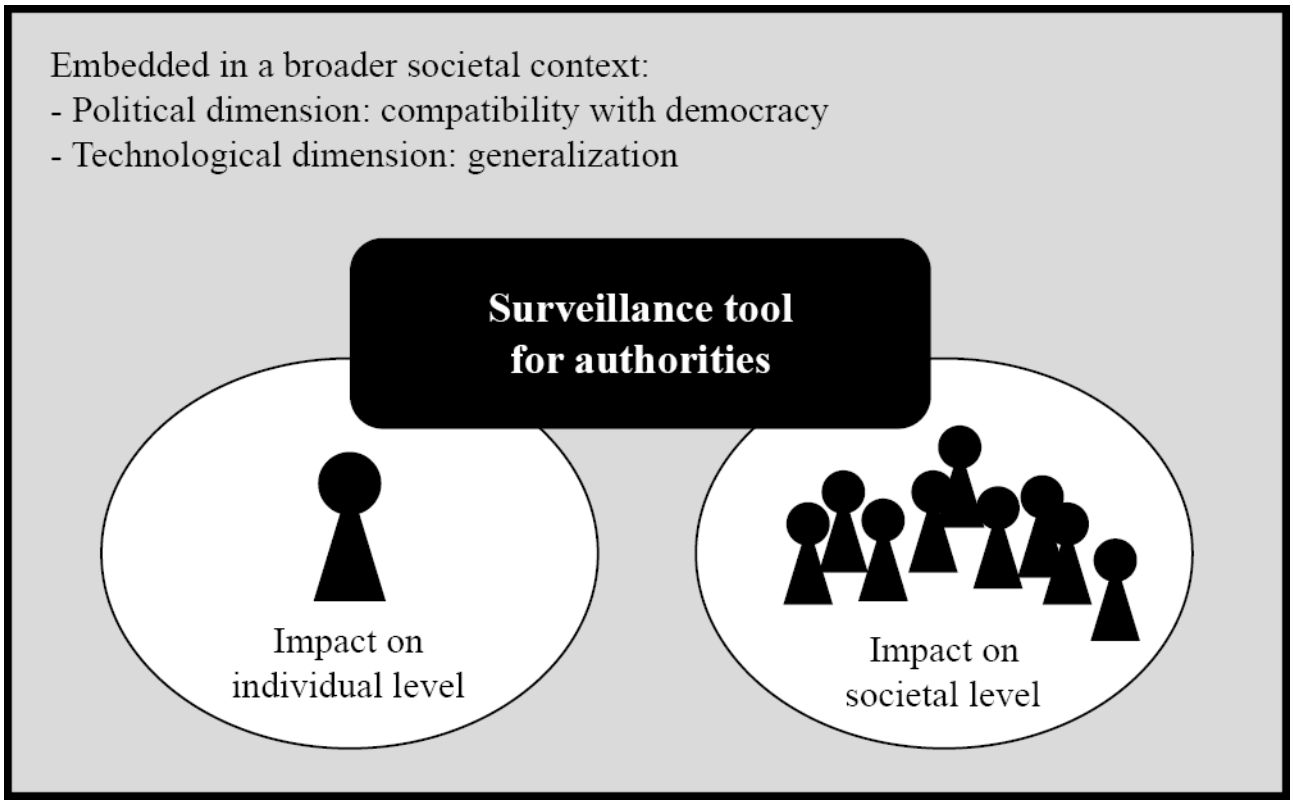
Perceiving Contact Tracing Apps as Governmental Surveillance Tools

Participants perceived contact tracing apps as governmental surveillance tools and embedded them in a broader societal context (Figure 2). However, “surveillance” was framed by participants in different ways. Some participants viewed such surveillance as a mechanism that empowers the authorities to control individual compliance with measures, which provoked negative sentiments. By contrast, other participants justified such contact tracing apps as surveillance tools for authorities that helped contain the pandemic without too many further restrictions. Yet other participants balanced the perils of surveillance against the specific needs to contain the pandemic:

> *Controlling is not necessarily a threat.* [German participant 7]

Although German and Swiss participants generally spoke about contract tracing apps only when asked about them specifically, approximately 20% of Austrian participants brought up the topic spontaneously, and their responses were also often more elaborate in length than the responses from Swiss and German participants.

Figure 2. Illustration of interview participants’ perceptions of digital contact tracing apps during the first wave of the COVID-19 pandemic (data collected in April and early May 2020).



Compatibility With Democracy

In all three countries, participants related contact tracing apps to Asian countries, such as China, which had already been using digital contact tracing technology for viral containment when the interviews were conducted. Although a few participants referred to the success of Asian countries in virus suppression as proof of the usefulness of these applications, most others used a comparison with “totalitarian states” as an argument against contact tracing apps, stating that the tools were incompatible with democratic values and rights.

> *Yes, I've heard about [the idea of contact tracing apps] and I completely reject it, I must say because my own data is simply too insecure, not protected and I don't want the state or whoever evaluates this data to know exactly where I was, what I did. And it also has a touch of the Chinese state or something [sic].*

> *In Asia, it is already being practiced without people critically questioning it. But I am absolutely against it.* [German participant 41]

By contrast, some participants framed contact tracing apps as tools that support authorities to control the viral spread:

> *I really believe in these apps because I think that if they are handled consistently, they can really reveal a lot of information to the authorities to develop certain measures or to react to the course of the disease [sic].* [German participant 3]

Participants also indicated a list of prerequisites that they considered crucial for a contact tracing app to be compatible with democracy: trust, privacy, voluntariness, and a time limit on data retention and use. The first prerequisite concerned the level of trust in national and local authorities. Since contact tracing apps were perceived as a surveillance measure of authorities, the level of trust towards these authorities was an important factor for people's willingness to participate. Participants in all three countries expressed the fear that the pandemic could be exploited by governments to install long-term surveillance systems, weighing the possible benefits of virus containment against the erosion of human rights:

> *I understand, of course, that this can be used to do contact tracing and potentially contain virus spread and I can see its usefulness. It's just, I'm struggling with the question of how much a crisis justifies intervening in basic human rights? And I think the problem is, if there is no pushback, then you might end up in a situation. I mean, I think that is what happened after the terrorist attacks in London and the US, that suddenly the NSAs of the world are created. And then nobody really knows what kind of information they actually have. And that shouldn't be the case in a democracy. So I think that citizens should be aware that some things are very difficult to undo later on.* [German participant 42]

The distrust of authorities was expressed most prominently by Austrian participants, who repeatedly called for more transparent communication about the app. They criticized a lack of clarity in communication surrounding how location and movement data were analyzed, possible extensions regarding the functionality of the app, or the role of the Austrian government in an app that was originally launched by Red Cross Austria, a nongovernmental organization. Austrian participants also seemed to be uncertain whether the app could be used for personalized tracking later on. They feared penalties for noncompliance or expressed concerns about a creeping loss of privacy. Others indicated skepticism surrounding a lack of clarity of how data would be used from the app.

> *So on the one hand the minister says: "Watch out because of this data on Facebook or whatever, or with people that want to call you or steal your login." On the other hand, he says we should use the Red Cross app. But they don't say what happens to the app or your data in the background [i.e. for other than contract tracing purposes], do they? I don't quite believe them, unfortunately. That's the only thing I don't really believe, to be honest.* [Austrian participant 46]

In the second prerequisite, many participants stressed the need to develop contact tracing apps that were compatible with existing privacy and data protection regulations:

> *I have heard about it [contact tracing apps] and I know that there are different models. So for me, it's just that I'm generally quite skeptical, that such things could be abused. For example, I don't do anything with YouTube or Facebook or whatever. I'm a bit old-fashioned in that sense. But last night I heard that someone in Switzerland has now created a program where data wouldn't go into any server anywhere, that it would just stay in your own smartphone and one wouldn't know any phone numbers of the people you met. And it wouldn't be stored anywhere. If that were so, I could say, yes, let's do it.* [Swiss participant 31]

Third, many participants believed that voluntariness was a key prerequisite for contact tracing apps to be compatible with democratic values and human rights. Although participants in all three countries mentioned this topic, we found it most predominantly mentioned in the Austrian data. Some Austrian participants even envisaged the possibility of "class action lawsuits" [Austrian participant 25] against the government and considered "strategies how to defend" themselves [Austrian participant 27] in case of the possibility of compulsory app use. Two Austrian participants even stated that they would rather die than be monitored by the government.

Finally, participants also thought that these tracing apps should be used temporary and only in the context of special circumstances such as the COVID-19 pandemic. Some participants suggested that the data should be erased when it was no longer needed, whereas others reiterated their concern that the data should only be used for the purposes indicated and not to build a surveillance state. One participant noted that although it made sense to use the app now, they intended to uninstall it by the end of the year, whereas another participant noted that these apps would no longer be justified when vaccines or other treatments are available.

Embedding Contact Tracing Apps in a General Technological Context

Participants from all countries linked tracing apps to other applications that collect data, such as social media, credit cards, or shopping cards, with the underlying concern that their data might be misused. The idea that others might be able to watch over one's actions and whereabouts was unnerving for some participants, particularly concerning uncertainty to where the data would be stored, who would have access to data for how long, and whether the crisis would erode privacy principles that would later be hard to re-establish:

> *I don't know, with technology, I always get the impression that these things creep in like crazy. Like we accept everything with every normal app because we only have the [binary] option not to use it or to accept the terms of use.* [Austrian participant 45]

Others were specifically concerned about the way the government or other powerful actors might use the information in the future:

> *We reveal so many things but we don't know yet how this could be used against us in the future […] So people should be careful and I believe that the state has far too much power over us in that context.* [Austrian participant 5]

Some participants seemed generally suspicious of surveillance technologies because of their potential for abuse, extending their suspicion of platforms such as YouTube or Facebook to digital tracking tools. Others said that their smartphones were already tools for individual surveillance. They suggested leaving smartphones at home to avoid being traced, or jokingly noted that they speak in dialect so that those "listening" will not understand:

> *I have a Huawei phone and I believe that the data will be delivered to China. That is why I always speak dialect because then the Chinese cannot understand me [laughing] [sic].* [Austrian participant 6]

Moreover, some participants worried that this technology could negatively affect not only their own lives but also how people relate to each other:

> *For me, it is not an option for everybody to be so high-tech. That's what it's all about. That you can see where somebody is. Who has been with whom in some way and so on. I find that threatening. I find that threatening. And I don't really want that.* [Swiss participant 22]

Embedding contact tracing apps in their general attitude towards technology led some to generally reject contact tracing apps. For others, by contrast, such consideration relativized their privacy concerns. For instance, one participant noted that they had already been recorded a hundred times on camera when driving into the city. Other participants suggested and expressed acceptance that people had long lost control over their data and their lives.

Impact on the Societal Level

Some participants framed contact tracing apps as a resource for the common good, stressing their function to help contain the viral spread and protect at-risk individuals. This was particularly pronounced in Germany and Switzerland, whereas only very few Austrian participants articulated this stance. One participant even felt that it was a duty to use the available technology to fight the risk of infection:

> *But I think that's good, I'm not concerned that [contact tracing apps restrict] freedom and privacy issues, instead I see the benefits. And in South Korea, [digital contact tracing] has really helped to reduce the number of infected people. And I believe that there is a duty to support this.* [German participant 35]

Others perceived tracing apps as a fast-track back to normality, increasing people's "freedom to move around" [German participant 26] and as a tool that "makes us all, as a society, more flexible" [Swiss participant 17]. By contrast, a few participants (particularly those in Austria) expressed fears of social pressure, stigma, and panic. One Austrian participant compared the app to the Star of David during the National Socialist Regime that singled out Jewish people. In a similar manner, this participant argued, the app stigmatizes infected people and makes them visible to everyone. It should be noted here that neither of the apps we report on here actually includes the function of making infected individuals visible to others.

Impact on the Individual Level

Some participants also focused on the impact the app might have on the individual level. Some of them focused on the potential personal benefit when using the app by framing contact tracing apps as a useful individual warning and information tools to proactively avoid COVID-19 infection. They hoped to get warnings to avoid risky situations, thereby decreasing personal uncertainty and providing orientation: "My wife and I, we would download and use it [contact tracing app] immediately and hope to be warned if someone comes too close. Or has come too close" [German participant 12]. This view was mainly taken by participants without practical experience with these apps, as none of the apps include such an immediate warning function. Moreover, contact tracing apps were framed as tools that affect individual responsibility. A few individuals perceived them as a tool that enables them to take individual responsibility: "I would probably also install the app on my cell phone, just because I find it helpful for me to know that if I picked up the virus somewhere, I could inform people, hey, watch out and stuff." [Austrian participant 15]. One participant said that they wished people would behave more responsibly, indicating that they felt contact tracing apps were only necessary because people were not acting responsibly enough.

Many of those participants who were skeptical of tracing apps were concerned about their privacy, expressing uneasiness about such tools "lying on the bedside table" [German participant 4] and comparing mobile phones to "personal diaries" [German participant 22] that they considered very private. Participants also expressed concerns about data protection in connection with contact tracing apps. To protect their data safely, some advocated for local data storage and against data silos.

Comparative Content Analysis of Newspaper Coverage

To further contextualize our participants' perceptions of contact tracing apps, we also performed a content analysis of newspaper coverage on contact tracing apps during the interview period (April 6 to May 6, 2020). By quantitatively and qualitatively comparing coverage from the four most-read newspapers of each country, we identified country-specific differences that laid out the basis for comparative interpretation as presented later in the discussion.

A total of 194 newspaper articles from 12 newspapers of the three German-speaking countries were included in the media content analysis (see Table 2). The nature of newspaper coverage regarding tracing apps was similar in the three countries studied, but Germany tended to have longer articles than the other two countries (see Multimedia Appendix 3).

Table 2. Sampling of newspapers and articles.

Country and newspaper	Genre	Publisher	Number of included articles
Switzerland			
Neue Zürcher Zeitung	International quality newspaper	NZZ Mediengruppe	31
Tages Anzeiger	National quality newspaper	Tamedia	19
Neue Luzerner Zeitung	Regional newspaper	CH Media	10
Blick	National tabloid	Ringier	5
Germany			
Süddeutsche Zeitung	International quality newspaper	Süddeutsche Zeitung GmbH	27
Die Welt	International quality newspaper	Axel Springer Verlag	23
taz, die tageszeitung	National quality newspaper	Taz Verlags-genossenschaft	16
BILD	National tabloid	Axel Springer Verlag	1
Austria			
Kurier	National quality newspaper	Kurier Zeitungsverlag und Druckerei GmbH	23
Der Standard	National quality newspaper	Standard Verlagsgesellschaft m. b. H.	16
Die Presse	National quality newspaper	Die Presse Verlags-Gesellschaft m.b.H. & Co KG	23
Krone.at	National tabloid	Mediaprint	0

Frames and Country-Specific Key Events

From the end of March to the beginning of April, German-language newspapers in all three countries reported geolocation or Wi-Fi–based tracking apps used for the anonymous analysis of population mobility or individual surveillance. Articles reported anonymous mobile phone data analysis to assess population mobility during the lockdown in all three countries. German and Austrian newspapers critically reported political discussions to use nonanonymous data for individual surveillance and raised fundamental concerns regarding basic human rights and democratic legitimacy. Swiss newspapers reported predominantly positive aspects and did not raise such concerns. In Austria, newspapers focused those concerns directly on the Austrian app, which was launched in March. In this context, the president of the first chamber of the Austrian Parliament, Wolfgang Sobotka, speculated publicly about making the Austrian app mandatory in an interview with the Austrian news magazine profil. This fueled already-existing concerns about civil rights and privacy violations in that context.

Similarly, the German health minister Jens Spahn attracted much critical coverage when, on March 25, 2020, he publicly considered including new options for tracking and surveillance in the updated German Epidemic Law. Even though newspapers in Germany reacted with similar concerns as those in Austria, the topic did not stay in the news as long as in Austria. Nevertheless, the launch of Robert Koch-Institut's "data donation app" on April 9, 2020, was accompanied by predominantly critical coverage about lack of transparency and insufficient data protection.

Starting at the end of March, the Swiss and German newspaper coverage discussed proximity tracing (meaning registering close contacts rather than using location data) as a possible assistance tool for contact tracing after the so-called lockdown period. Even though data protection and privacy issues were discussed, proximity tracing was introduced to potentially overcome these issues. To that end, the German and Swiss newspapers evaluated such applications more positively than Austrian newspapers, which were predominantly critical towards the potential privacy implications of the Austrian app. In mid-April, German and Swiss media picked up disagreements on data storage mechanisms within the PEPP-PT research consortium. Swiss coverage focused on the Swiss epidemiologist Marcel Salathé who argued for a decentralized data storage system. In that context, Swiss newspapers reported supportively about the Swiss contact tracing app in development. In Germany, where researchers had initially favored a centralized data storage system, the same éclat caused discussions about data protection requirements. On April 27, 2020, the German government announced that they had decided on a decentralized data storage system. This was evaluated positively in the newspapers. All three countries also compared nationally discussed solutions of contact tracing apps with international applications, especially those in active use in Israel, China, and South Korea, and criticized those as being largely inconceivable in Western European democracies. Country-specific framings and key events are presented in detail in Multimedia Appendix 4.

National and International Views on Coverage

In Swiss and German coverage, the international collaboration of researchers involved in research needed for the development of such an app was mentioned more often than in Austria (see Table 3).

Table 3. Stakeholders cited in newspaper articles.

Stakeholders cited	Mentions, n (%)		
	Germany (n=67)	Austria (n=62)	Switzerland (n=65)
Governmental actors or politicians	28 (42)	30 (48)	36 (55)
Nongovernmental actors	9 (13)	11 (18)	1 (2)
Scientific and medical experts	22 (33)	5 (8)	32 (49)
Legal experts	4 (6)	18 (29)	9 (14)
Experts from humanities or social sciences	9 (13)	1 (2)	2 (3)
Celebrities or VIPs	1 (1)	1 (2)	0 (0)
Civil society	2 (3)	0 (0)	2 (3)
Private companies	4 (6)	3 (5)	4 (6)
Other	0 (0)	1 (2)	0 (0)

In general, Austrian coverage tended to focus more on a national perspective than Swiss and German coverage, where newspapers repeatedly referred to the international PEPP-PT research initiative. This, in turn, led to coverage on the topic of centralized versus decentralized data storage in these two countries, which did not come up as frequently in Austrian newspaper coverage (see Table 4). In all three countries, tracing and tracking applications were repeatedly connected to totalitarian surveillance states, such as China, or even Asian democracies such as South Korea, where surveillance apps were used as containment strategies.

Table 4. Topics mentioned in newspaper articles.

Topic	Germany (n=67)		Austria (n=62)		Switzerland (n=65)	
	Topic total, n (%)	Main topic[a], n (%)	Topic total, n (%)	Main topic, n (%)	Topic total, n (%)	Main topic, n (%)
GPS motion tracking at population level (anonymized)	3 (4)	2 (3)	4 (6)	0 (0)	16 (25)	6 (9)
GPS motion tracking on an individual level	12 (18)	4 (6)	13 (21)	6 (10)	4 (6)	1 (2)
“Datenspende” app (Robert Koch-Institut)	8 (12)	5 (7)	0 (0)	0 (0)	0 (0)	0 (0)
Development of contact tracing apps	14 (21)	10 (15)	20 (32)	11 (18)	26 (40)	9 (14)
Functioning of contact tracing apps	19 (28)	9 (13)	18 (29)	6 (10)	24 (37)	4 (6)
Centralized vs decentralized data storage	12 (18)	4 (6)	1 (2)	1 (2)	7 (11)	0 (0)
Legal or ethical issues regarding tracing apps	5 (7)	3 (4)	38 (61)	19 (31)	33 (51)	15 (23)
Other relevant topics	4 (6)	4 (6)	0 (0)	0 (0)	0 (0)	0 (0)
Tracking/tracing technology that not the main topic	25 (37)	N/A[b]	17 (27)	N/A	29 (45)	N/A

[a]A topic was defined as the “main topic” if it was mentioned in the title or the first paragraph of the article. There was only one main topic per article but otherwise an indefinite number of topics could be coded if necessary.

[b]Not applicable.

Discussion

Even though policymakers in Germany, Austria, and Switzerland framed contact tracing apps as safe and helpful tools to contain the viral spread, most participants perceived contact tracing apps as governmental surveillance tools. This perception was represented in all interviews and across all three countries. What varied, however, was how people framed and assessed this form of governmental surveillance. Many participants reflected on the compatibility of contact tracing apps with Western democracies. Newspaper coverage periodically framed these apps in a similar fashion by using three frames that were predominantly covered in April 2020: (1) critical examination of the impact of contact tracing apps on individual privacy and the compatibility with existing data protection regulations; (2) simultaneous framing and comparison with digital surveillance tools used in Asian countries, especially China, provoking a contrasting view to Western democracies; and (3) periodical focus on country-specific political and scientific stakeholders.

Privacy issues were a predominant concern expressed by participants and in newspaper coverage; these issues also mirror the early concerns expressed in the literature [43,60]. People were concerned about the impact of contact tracing apps on individual freedom and privacy, in line with concerns commonly

expressed in the newspaper articles reviewed in this study. Participants emphasized that it was a matter of principle for them: The use of a digital tool that was perceived to be so intrusive in people's privacy should never be made mandatory. Indeed, the voluntariness of contact tracing apps is also debated and emphasized by experts, including what voluntariness means in this context [61-63]. Nevertheless, ethicists have more recently criticized this emphasis on privacy and pointed to other ethical issues that were left aside in app development, such as safety and effectiveness [60,64], social justice issues [65,66], or the potentially problematic policy influence of Apple and Google in that context [18,67].

Possible Explanations for Low Uptake Rates

At the time of writing this manuscript, over half a year after the launch of the respective contact tracing apps, app developers and authorities were facing lower user rates than initially expected [8,9]. This constitutes some tension with other government measures that also limited people's freedom but were readily accepted; for example, the majority of people in Germany and Austria supported compulsory face mask use a few months later [12,68,69].

One reason for this might be that the early perception as a surveillance tool for authorities left many people reluctant to use contact tracing apps. The observed tendency to put these contact tracing apps into a broader context of technological surveillance tools and privacy concerns might have triggered general concerns of privacy that are known to be particularly important in the German-speaking areas [70], and the same has been reported in the French context [18]. Newspapers reporting the use of contact tracing and tracking apps in Asian countries might have further deepened the impression that these applications are not aligned with democratic principles of individual privacy and freedom since these applications were considered by interview participants and newspapers alike to be not compatible with privacy regulations and democratic principles as understood in the European context.

The broad contextualization of contact tracing apps makes factual information about the app itself (such as it contributes to controlling the pandemic, data is anonymized, and location tracking impossible) less effective, as people tend to use pre-existing concepts from other technologies to avoid cognitive dissonance, which occurs when expectations and performance contradict each other [71]. Thus, even though the performances of the apps are privacy-friendly, those who are generally skeptical towards digital tools and surveillance still expect them to be dangerous, despite the crucial efforts of app developers and policymakers. This points to the underlying and unsolved issues of regulating digital tools, data ownership, and nontransparent use and economization of data [67].

Likewise, Barth and Jong [72] suggested that the privacy paradox, in which people share some information willingly but are reluctant to share information with others, was relying on risk-benefit assessment (see also [60]). It is plausible that a subset of people considers contact tracing apps to have more privacy-related risks than benefits. Indeed, the currently available contact tracing apps have little personal benefit but promise to serve the common good by contributing to pandemic containment. Proof of effectiveness has been difficult to demonstrate thus far because user rates have been lower than originally suggested [14], so potential benefits could not unfold, and it is challenging to collect relevant data for quality control due to privacy constraints. Showing the benefits of contact tracing apps on an individual and a societal level might increase the uptake, but would need to be reliable, ideally through direct empirical evidence of effectiveness.

Public Trust and Transparent Communication

Particularly in Austrian interviews, people criticized what they perceived to be the absence of transparent political communication. For example, the public speculation of the president of the first chamber of the Austrian parliament, Wolfgang Sobotka, that the use of the Austrian app may become mandatory seemed to undermine people's trust. It also influenced participants' responses in the days following this particular event. Although most Austrian participants trusted the Red Cross, the fact that Red Cross launched the app was often considered less relevant than the function of the app as a surveillance tool for authorities. Trustworthiness was generally a less dominant topic in Swiss and German interviews. In both countries, epidemiologists and app developers were frequently given a voice in newspapers. For instance, newspaper coverage followed discussions of app developers about centralized or decentralized data storage, or it critically reflected on political considerations connected to voluntariness and whether to introduce a legal basis for surveillance measures. These issues were publicly discussed and addressed prior to the release of contact tracing apps in June, and upon their launch, it was already legally binding that apps could not become compulsory and needed to employ decentralized storage systems. This is in sharp contrast with other examples, such as in France, where a centralized storage system was implemented despite privacy concerns [18]. Such discussions prior to launch might have promoted public trust in authorities and could be a possible reason for the higher relative uptake numbers in Germany and Switzerland as compared with those in Austria. This is further supported by our finding that German and Swiss newspapers framed the discussion, particularly the scientific discussion behind app development, from a more international perspective, whereas Austrian newspapers focused on the national issues of the already-released app. In line with our findings, recent ethical inquiries have also identified public trust as the most critical element for the uptake of contact tracing apps [18,40,73-75].

Conflation of Different Mobile Phone–Based Tracing and Tracking Apps

Newspapers and interview participants alike often conflated different tracking applications (anonymous geolocation or Wi-Fi data population analysis vs. real-time personal GPS tracking versus Bluetooth-based proximity tracing). This seems to have confused people and may have solidified concerns surrounding the possibility of ubiquitous surveillance because instances where de-identified information was collected were interpreted as personalized tracking. These uncertainties and misconceptions seemed to be associated with uneasiness and general skepticism. These findings highlight the difficult task for governments to communicate about the apps they plan to introduce transparently,

clearly, and simply, while including sufficient information to allow people to distinguish between different applications [76,77].

Limitations

This study did not assess the exact prevalence of specific views on, or experiences with, contact tracing apps in the general population in the three countries or regions examined. Instead, we sought to explore how people describe their practices regarding contract tracing apps where they have had personal experience, and their normative, factual, and emotional reference points when discussing the design, utility, and effect of digital contact tracing. To contextualize the data obtained from the large-scale interview study, we also carried out an analysis of mainstream news media in the three countries or regions that was limited by the fact that newspapers were the only format examined and other mass media (such as television, radio, and social media) were excluded. No reliability testing was conducted because we interpreted the data relationally and qualitatively, looking for relationships rather than statistical significance. As such, this article provides context-specific insights into digital tracking tools during a specific period in the COVID-19 pandemic. Even though we sought a balanced demographic distribution, the final sample of interview participants is slightly skewed towards people with higher educational levels (more than 10 years), those who are under 70 years of age, and those in the Austrian sample living in urban areas. However, because this is a qualitative analysis not aiming for statistical representativeness, every view was considered independent of how many participants had stated this view. Given that several interviewees supported every view we report and that we were able to draw what in the grounded theory approach is called a "middle-range theory," we conclude that we have reached satisfying theoretical saturation [78]. Although we did report when issues were mentioned particularly frequently or particularly rarely, any quantification in terms of how many citizens support these views is subject to further quantitative inquiries, such as follow-up surveys, which go beyond the scope of this paper. In the Swiss context, only the German-speaking part is covered both in terms of newspaper coverage and interview participants. Since there are considerable cultural differences between Swiss language regions [79], our findings are limited to the German-speaking part of Switzerland.

Conclusions

Newspapers from and people living in German-speaking countries perceived digital contact tracing apps as surveillance tools popularized by authorities during the first wave of the COVID-19 pandemic in April 2020, attributing a broad range of interpretations, evaluations, and impact to this perception. Even though privacy was a common concern among participants, many also framed surveillance in a positive way and saw contact tracing apps as tools that could benefit society in containing the viral spread. Others saw the potential for personal and societal benefit, affording users better control over their exposure risk. Voluntariness and trustworthiness were most frequently discussed by Austrian participants, in line with Austrian newspaper coverage and political discussions at that time. The early launch of the Austrian Stopp Corona app has, on the one hand, made the topic more immediately relevant for both study participants and mass media. On the other hand, ongoing debates about the voluntariness and the use of apps for checking individual compliance through the authorities might have caused more severe and lasting rejections in Austria than in Germany or Switzerland, where the relative uptake of the apps was slightly higher.

Although communication from authorities and app developers shapes peoples' early perceptions of contact tracing apps, their previous experiences and expectations with authorities and digital tools also play an important role. Thus, when authorities plan to implement new digital tools and practices in the future, they should be very clear in communicating the objectives, the contribution of this technology, and how it differs from other, possibly similar, tools that may be problematic or may have been previously used in a problematic way. Similarly, even in cases where there is a pressing need for the use of new tools, it is important to address publicly and solve ethical, legal, and social issues related to such technologies prior to their launch—existing concerns related to new technologies need to be addressed proactively. Finally, to overcome the privacy paradox, the effectiveness of contact tracing apps needs to be evaluated and communicated.

Acknowledgments

The interview study was conducted as part of the SolPan consortium, and we acknowledge all consortium members for their valuable discussions and contributions concerning study design and strategy. We particularly thank Dr. Katharina Kieslich for her insights into Austrian contact tracing app policies and student assistants Magnus Tibbe and Justus Bredthauer for extracting data from Austrian and Swiss newspaper articles. This work was supported by the German Federal Ministry for Education and Research (Bundesministerium für Bildung und Forschung; Grant number 01Kl20510), which had no role in the project design; in the collection, analysis, or interpretation of data; in the writing of the article; or in the decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Qualitative interview guide (SolPan; translated German version, April 2020).

[DOCX File , 24 KB - jmir_v23i2e25525_app1.docx]

Multimedia Appendix 2
Codebook used for newspaper content analysis.
[DOCX File , 29 KB - jmir_v23i2e25525_app2.docx]

Multimedia Appendix 3
Supplementary tables of newspaper content analysis.
[DOCX File , 13 KB - jmir_v23i2e25525_app3.docx]

Multimedia Appendix 4
Country-specific analysis of frames portrayed in newspaper coverage over time.
[DOCX File , 24 KB - jmir_v23i2e25525_app4.docx]

References

1. Die Corona-Warn-App: Unterstützt uns im Kampf gegen Corona. German Federal Government. 2021. URL: https://www.bundesregierung.de/breg-de/themen/corona-warn-app [accessed 2021-01-11]
2. SwissCovid App. Swiss Federal Office of Public Health. 2021. URL: https://bag-coronavirus.ch/swisscovid-app/ [accessed 2021-01-11]
3. Questions and answers about the Stop Corona app. Stop Corona App (Austrian Red Cross). 2021. URL: https://www.stopp-corona.at/faq_stopp_corona_app/ [accessed 2021-01-11]
4. PEPP-PT Documentation. PEPP-PT Consortium. URL: https://github.com/pepp-pt/pepp-pt-documentation [accessed 2021-01-11]
5. DP-3T Documents. DP-3T Consortium. URL: https://github.com/DP-3T/documents [accessed 2021-01-11]
6. Grube T, Heinrich A, Stroscher J, Schomberg S. Data security of Corona apps according to the GDPR (Article in German). Data Protection Data Security 2020 Jul 27;44(8):501-505. [doi: 10.1007/s11623-020-1314-0]
7. Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, et al. A Survey of COVID-19 contact tracing apps. IEEE Access 2020;8:134577-134601. [doi: 10.1109/access.2020.3010226]
8. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
9. Bosshardt L, Bühler G, Craviolini J, Hermann M, Krähenbühl D. Federal Tracing App (Report in German). Zurich: Federal Office of Public Health; 2020 May 25. URL: https://www.newsd.admin.ch/newsd/message/attachments/61463.pdf [accessed 2021-01-28]
10. Key figures for the Corona-Warn-App (Article in German). Robert Koch Institute. 2020 Oct 30. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Archiv_Kennzahlen/Kennzahlen_30102020.pdf?__blob=publicationFile [accessed 2020-11-04]
11. Partheymüller J, Kritzinger S, Song H, Plescia C. Learn from South Korea? For the use and acceptance of the “Stop Corona” app in Austria (Article in German). University of Vienna - Corona Blog 31. 2020 May 06. URL: https://viecer.univie.ac.at/corona-blog/corona-blog-beitraege/blog31/ [accessed 2021-01-11]
12. Kittel B, Kritzinger S, Boomgaarden H, Prainsack B, Eberl J, Kalleitner F, et al. The Austrian Corona Panel Project: monitoring individual and societal dynamics amidst the COVID-19 crisis. SSRN Journal. Preprint posted online on July 23, 2020 2020 [FREE Full text] [doi: 10.2139/ssrn.3654139]
13. Aichholzer J, Kalleitner F. “Stop Corona” app: development in use and acceptance (Article in German). University of Vienna - Corona Blog. 2020 Nov 13. URL: https://viecer.univie.ac.at/corona-blog/corona-blog-beitraege/corona-dynamiken9/ [accessed 2021-01-11]
14. Hinch R, Probert W, Nurtay A, Kendall M, Wymant C, Hall M, et al. Effective Configurations of a Digital Contact Tracing App: A report to NHSX. 2020 Apr 16. URL: https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf?1587531217 [accessed 2021-02-02]
15. Hernández-Orallo E, Calafate CT, Cano J, Manzoni P. Evaluating the effectiveness of COVID-19 Bluetooth-based smartphone contact tracing applications. Applied Sciences 2020 Oct 13;10(20):7113. [doi: 10.3390/app10207113]
16. Salathé M, Althaus C, Anderegg N, Antonioli D, Ballouz T, Bugnon E, et al. Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland. Swiss Med Wkly 2020 Dec 14;150:w20457 [FREE Full text] [doi: 10.4414/smw.2020.20457] [Medline: 33327003]
17. Maccari L, Cagno V. Do we need a contact tracing app? Comput Commun 2021 Jan 15;166:9-18 [FREE Full text] [doi: 10.1016/j.comcom.2020.11.007] [Medline: 33235399]
18. Rowe F, Ngwenyama O, Richet J. Contact-tracing apps and alienation in the age of COVID-19. European Journal of Information Systems 2020 Sep 13;29(5):545-562. [doi: 10.1080/0960085x.2020.1803155]

19. Mbunge E. Integrating emerging technologies into COVID-19 contact tracing: opportunities, challenges and pitfalls. Diabetes Metab Syndr 2020;14(6):1631-1636. [doi: 10.1016/j.dsx.2020.08.029] [Medline: 32892060]
20. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
21. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR mHealth uHealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
22. Birkhoff SD, Smeltzer SC. Perceptions of smartphone user-centered mobile health tracking apps across various chronic illness populations: an integrative review. J Nurs Scholarsh 2017 Jul;49(4):371-378. [doi: 10.1111/jnu.12298] [Medline: 28605151]
23. Deng Z, Mo X, Liu S. Comparison of the middle-aged and older users' adoption of mobile health services in China. Int J Med Inform 2014 Mar;83(3):210-224. [doi: 10.1016/j.ijmedinf.2013.12.002] [Medline: 24388129]
24. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
25. Kwon M, Mun K, Lee JK, McLeod DM, D'Angelo J. Is mobile health all peer pressure? The influence of mass media exposure on the motivation to use mobile health apps. Convergence 2016 Apr 10;23(6):565-586. [doi: 10.1177/1354856516641065]
26. Simblett S, Matcham F, Siddi S, Bulgari V, Barattieri di San Pietro C, Hortas López J, RADAR-CNS Consortium. Barriers to and facilitators of engagement with mHealth technology for remote measurement and management of depression: qualitative analysis. JMIR mHealth uHealth 2019 Jan 30;7(1):e11325 [FREE Full text] [doi: 10.2196/11325] [Medline: 30698535]
27. Spann A, Stewart E. Barriers and facilitators of older people's mHealth usage: a qualitative review of older people's views. Human Technology 2018 Nov 30:264-296. [doi: 10.17011/ht/urn.201811224834]
28. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
29. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han H. Facilitators of and barriers to mHealth adoption in older adults with heart failure. Comput Inform Nurs 2018 Aug;36(8):376-382 [FREE Full text] [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
30. Stach M, Kraft R, Probst T, Messner EM, Terhorst Y, Baumeister H, et al. Mobile health app database - a repository for quality ratings of mHealth apps. : IEEE; 2020 Presented at: IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS); July 28-30, 2020; Rochester, MN p. 427-432. [doi: 10.1109/CBMS49503.2020.00087]
31. Korpershoek YJG, Vervoort SCJM, Trappenburg JCA, Schuurmans MJ. Perceptions of patients with chronic obstructive pulmonary disease and their health care providers towards using mHealth for self-management of exacerbations: a qualitative study. BMC Health Serv Res 2018 Oct 04;18(1):757 [FREE Full text] [doi: 10.1186/s12913-018-3545-4] [Medline: 30286761]
32. Sheats JL, Petrin C, Darensbourg RM, Wheeler CS. A Theoretically-Grounded Investigation of Perceptions About Healthy Eating and mHealth Support Among African American Men and Women in New Orleans, Louisiana. Fam Community Health 2018;41 Suppl 2 Suppl, Food Insecurity and Obesity:S15-S24 [FREE Full text] [doi: 10.1097/FCH.0000000000000177] [Medline: 29461312]
33. Dallinga J, Janssen M, van der Werf J, Walravens R, Vos S, Deutekom M. Analysis of the Features Important for the Effectiveness of Physical Activity-Related Apps for Recreational Sports: Expert Panel Approach. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e143 [FREE Full text] [doi: 10.2196/mhealth.9459] [Medline: 29914863]
34. Trang S, Trenz M, Weiger W, Tarafdar M, Cheung C. One app to trace them all? Examining app specifications for mass acceptance of contact-tracing apps. European Journal of Information Systems 2020 Jul 27;29(4):415-428. [doi: 10.1080/0960085X.2020.1784046]
35. Guillon M, Kergall P. Attitudes and opinions on quarantine and support for a contact-tracing application in France during the COVID-19 outbreak. Public Health 2020 Nov;188:21-31 [FREE Full text] [doi: 10.1016/j.puhe.2020.08.026] [Medline: 33059232]
36. Joo J, Shin M. Resolving the tension between full utilization of contact tracing app services and user stress as an effort to control the COVID-19 pandemic. Serv Bus 2020 Sep 01;14(4):461-478 [FREE Full text] [doi: 10.1007/s11628-020-00424-7] [Medline: 18300068]
37. O'Callaghan ME, Buckley J, Fitzgerald B, Johnson K, Laffey J, McNicholas B, et al. A national survey of attitudes to COVID-19 digital contact tracing in the Republic of Ireland. Ir J Med Sci 2020 Oct 16 [FREE Full text] [doi: 10.1007/s11845-020-02389-y] [Medline: 33063226]
38. Sharma S, Singh G, Sharma R, Jones P, Kraus S, Dwivedi Y. Digital Health Innovation: Exploring Adoption of COVID-19 Digital Contact Tracing Apps. IEEE Trans. Eng. Manage 2004:1-17. [doi: 10.1109/TEM.2020.3019033]

39. Walrave M, Waeterloos C, Ponnet K. Ready or Not for Contact Tracing? Investigating the Adoption Intention of COVID-19 Contact-Tracing Technology Using an Extended Unified Theory of Acceptance and Use of Technology Model. Cyberpsychol Behav Soc Netw 2020 Oct 05. [doi: 10.1089/cyber.2020.0483] [Medline: 33017171]
40. Kaspar K. Motivations for social distancing and app use as complementary measures to combat the COVID-19 pandemic: quantitative survey study. J Med Internet Res 2020 Aug 27;22(8):e21613 [FREE Full text] [doi: 10.2196/21613] [Medline: 32759100]
41. Keshet Y. Fear of panoptic surveillance: using digital technology to control the COVID-19 epidemic. Isr J Health Policy Res 2020 Nov 25;9(1):67 [FREE Full text] [doi: 10.1186/s13584-020-00429-7] [Medline: 33239094]
42. Eck K, Hatz S. State surveillance and the COVID-19 crisis. Journal of Human Rights 2020 Nov 11;19(5):603-612 [FREE Full text] [doi: 10.1080/14754835.2020.1816163] [Medline: 32836636]
43. Fahey RA, Hino A. COVID-19, digital privacy, and the social limits on data-focused public health responses. Int J Inf Manage 2020 Dec;55:102181 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102181] [Medline: 32836638]
44. Zhang M, Chow A, Smith H. COVID-19 Contact-Tracing Apps: Analysis of the Readability of Privacy Policies. J Med Internet Res 2020 Dec 03;22(12):e21572 [FREE Full text] [doi: 10.2196/21572] [Medline: 33170798]
45. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):e28-e33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
46. Nikolaou CK, Tay Z, Leu J, Rebello SA, Te Morenga L, Van Dam RM, et al. Young People's Attitudes and Motivations Toward Social Media and Mobile Apps for Weight Control: Mixed Methods Study. JMIR Mhealth Uhealth 2019 Oct 10;7(10):e11205 [FREE Full text] [doi: 10.2196/11205] [Medline: 31603431]
47. Willcox JC, van der Pligt P, Ball K, Wilkinson SA, Lappas M, McCarthy EA, et al. Views of Women and Health Professionals on mHealth Lifestyle Interventions in Pregnancy: A Qualitative Investigation. JMIR Mhealth Uhealth 2015 Oct 28;3(4):e99 [FREE Full text] [doi: 10.2196/mhealth.4869] [Medline: 26510886]
48. Alexopoulos AR, Hudson JG, Otenigbagbe O. The Use of Digital Applications and COVID-19. Community Ment Health J 2020 Oct;56(7):1202-1203 [FREE Full text] [doi: 10.1007/s10597-020-00689-2] [Medline: 32734311]
49. Soroya SH, Farooq A, Mahmood K, Isoaho J, Zara S. From information seeking to information avoidance: Understanding the health information behavior during a global health crisis. Inf Process Manag 2021 Mar;58(2):102440 [FREE Full text] [doi: 10.1016/j.ipm.2020.102440] [Medline: 33281273]
50. Basch CH, Kecojevic A, Wagner VH. Coverage of the COVID-19 pandemic in the online versions of highly circulated U.S. daily newspapers. J Community Health 2020 Dec;45(6):1089-1097 [FREE Full text] [doi: 10.1007/s10900-020-00913-w] [Medline: 32902813]
51. Zimmermann BM, Aebi N, Kolb S, Shaw D, Elger BS. Content, evaluations and influences in newspaper coverage of predictive genetic testing: A comparative media content analysis from the United Kingdom and Switzerland. Public Underst Sci 2019 Apr;28(3):256-274. [doi: 10.1177/0963662518816014] [Medline: 30583711]
52. Donelle L, Hoffman-Goetz L, Clarke JN. Ethnicity, genetics, and breast cancer: media portrayal of disease identities. Ethn Health 2005 Aug;10(3):185-197. [doi: 10.1080/13557850500120751] [Medline: 16087452]
53. Niederdeppe J, Frosch DL, Hornik RC. Cancer news coverage and information seeking. J Health Commun 2008 Mar;13(2):181-199 [FREE Full text] [doi: 10.1080/10810730701854110] [Medline: 18300068]
54. Pierce JP, Gilpin EA. News media coverage of smoking and health is associated with changes in population rates of smoking cessation but not initiation. Tob Control 2001 Jun;10(2):145-153 [FREE Full text] [doi: 10.1136/tc.10.2.145] [Medline: 11387535]
55. Wirth W, Kolb S. Designs and methods of comparative political communication research. In: Esser F, Pfetsch B, editors. Comparing Political Communication: Theories, Cases, and Challenges (Communication, Society and Politics). Cambridge: Cambridge University Press; 2004:87-112.
56. Solidarity in times of a pandemic: What do people do, and why? University of Vienna - Research Platform: Governance of digital practices. URL: https://digigov.univie.ac.at/solidarity-in-times-of-a-pandemic-solpan/ [accessed 2021-01-11]
57. Wirth W, Kolb S. Securing equivalence: problems and solutions. In: Esser F, Hanitzsch T, editors. The Handbook of Comparative Communication Research (1st Edition). Abingdon: Routledge; Mar 2012:469-485.
58. Su Y, Borah P. Who is the agenda setter? Examining the intermedia agenda-setting effect between Twitter and newspapers. Journal of Information Technology & Politics 2019 Jul 13;16(3):236-249. [doi: 10.1080/19331681.2019.1641451]
59. Dow Jones Professional. URL: https://professional.dowjones.com/factiva/ [accessed 2021-02-02]
60. Rowe F. Contact tracing apps and values dilemmas: A privacy paradox in a neo-liberal world. Int J Inf Manage 2020 Dec;55:102178 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102178] [Medline: 32836636]
61. Bengio Y, Janda R, Yu YW, Ippolito D, Jarvie M, Pilat D, et al. The need for privacy with public digital contact tracing during the COVID-19 pandemic. Lancet Digit Health 2020 Jul;2(7):e342-e344 [FREE Full text] [doi: 10.1016/S2589-7500(20)30133-3] [Medline: 32835192]
62. Loi M. How to fairly incentivise digital contact tracing. J Med Ethics 2020 Jul 09 [FREE Full text] [doi: 10.1136/medethics-2020-106388] [Medline: 32647047]
63. Siffels LE. Beyond privacy vs. health: a justification analysis of the contact-tracing apps debate in the Netherlands. Ethics Inf Technol 2020 Oct 10:1-5 [FREE Full text] [doi: 10.1007/s10676-020-09555-x] [Medline: 33071607]

64. Martinez-Martin N, Wieten S, Magnus D, Cho MK. Digital contact tracing, privacy, and public health. Hastings Cent Rep 2020 May;50(3):43-46 [FREE Full text] [doi: 10.1002/hast.1131] [Medline: 32596893]
65. Hendl T, Chung R, Wild V. Pandemic surveillance and racialized subpopulations: mitigating vulnerabilities in COVID-19 apps. J Bioeth Inq 2020 Dec;17(4):829-834 [FREE Full text] [doi: 10.1007/s11673-020-10034-7] [Medline: 32840858]
66. Klenk M, Duijf H. Ethics of digital contact tracing and COVID-19: who is (not) free to go? Ethics Inf Technol 2020 Aug 24:1-9 [FREE Full text] [doi: 10.1007/s10676-020-09544-0] [Medline: 32863740]
67. Sharon T. Blind-sided by privacy? Digital contact tracing, the Apple/Google API and big tech's newfound role as global health policy makers. Ethics Inf Technol 2020 Jul 18:1-13 [FREE Full text] [doi: 10.1007/s10676-020-09547-x] [Medline: 32837287]
68. Kalleitner F, Pollak M, Partheymueller J. High approval of the mask requirement: 81% think that it should apply (Article in German). University of Vienna - Corona Blog. 2020 Sep 28. URL: https://viecer.univie.ac.at/corona-blog/corona-blog-beitraege/corona-dynamiken2/ [accessed 2021-01-11]
69. Betsch C, Korn L, Sprengholz P, Felgendreff L, Eitze S, Schmid P, et al. Social and behavioral consequences of mask policies during the COVID-19 pandemic. Proc Natl Acad Sci U S A 2020 Sep 08;117(36):21851-21853 [FREE Full text] [doi: 10.1073/pnas.2011674117] [Medline: 32820078]
70. Voigt TH, Holtz V, Niemiec E, Howard HC, Middleton A, Prainsack B. Willingness to donate genomic and other medical data: results from Germany. Eur J Hum Genet 2020 Aug;28(8):1000-1009 [FREE Full text] [doi: 10.1038/s41431-020-0611-2] [Medline: 32238912]
71. Marikyan D, Papagiannidis S, Alamanos E. Cognitive dissonance in technology adoption: a study of smart home users. Inf Syst Front 2020 Jul 25:1-23 [FREE Full text] [doi: 10.1007/s10796-020-10042-3] [Medline: 32837263]
72. Barth S, de Jong MD. The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – a systematic literature review. Telematics and Informatics 2017 Nov;34(7):1038-1058. [doi: 10.1016/j.tele.2017.04.013]
73. Ranisch R, Nijsingh N, Ballantyne A, van Bergen A, Buyx A, Friedrich O, et al. Digital contact tracing and exposure notification: ethical guidance for trustworthy pandemic management. Ethics Inf Technol 2020 Oct 21:1-10 [FREE Full text] [doi: 10.1007/s10676-020-09566-8] [Medline: 33106749]
74. The app credibility gap. Nat Biotechnol 2020 Jul;38(7):768 [FREE Full text] [doi: 10.1038/s41587-020-0610-4] [Medline: 32591763]
75. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020 Jul;46(7):427-431 [FREE Full text] [doi: 10.1136/medethics-2020-106314] [Medline: 32366705]
76. Coghlan S, Cheong M, Coghlan B. Tracking, tracing, trust: contemplating mitigating the impact of COVID-19 through technological interventions. Med J Aust 2020 Jul;213(2):94-94.e1 [FREE Full text] [doi: 10.5694/mja2.50680] [Medline: 32570292]
77. Lucivero F, Hallowell N, Johnson S, Prainsack B, Samuel G, Sharon T. COVID-19 and contact tracing apps: ethical challenges for a social experiment on a global scale. J Bioeth Inq 2020 Dec;17(4):835-839 [FREE Full text] [doi: 10.1007/s11673-020-10016-9] [Medline: 32840842]
78. Low J. A pragmatic definition of the concept of theoretical saturation. Sociological Focus 2019 Jan 27;52(2):131-139. [doi: 10.1080/00380237.2018.1544514]
79. Deopa N, Forunato P. Coronagraben. Culture and social distancing in times of COVID-19. SSRN Journal. Preprint posted online on Jun 29, 2020 2020. [doi: 10.2139/ssrn.3635287]

Abbreviations

DP-3T: Decentralized Privacy-Preserving Proximity Tracing
mHealth: mobile health
PEPP-PT: Pan-European Privacy-Preserving Proximity Tracing
SolPan: Solidarity in Times of a Pandemic

Edited by G Eysenbach; submitted 05.11.20; peer-reviewed by F Rowe, A Chang, M Ariyan, Z Aghaei, A Khaleghi; comments to author 02.12.20; revised version received 17.12.20; accepted 09.01.21; published 08.02.21.

Please cite as:
Zimmermann BM, Fiske A, Prainsack B, Hangel N, McLennan S, Buyx A
Early Perceptions of COVID-19 Contact Tracing Apps in German-Speaking Countries: Comparative Mixed Methods Study
J Med Internet Res 2021;23(2):e25525
URL: http://www.jmir.org/2021/2/e25525/
doi:10.2196/25525
PMID:33503000

Original Paper

The Role of Transparency, Trust, and Social Influence on Uncertainty Reduction in Times of Pandemics: Empirical Study on the Adoption of COVID-19 Tracing Apps

Andreas Oldeweme[1], MSc; Julian Märtins[1], MSc; Daniel Westmattelmann[1], PhD; Gerhard Schewe[1], PhD

University of Muenster, Muenster, Germany

Corresponding Author:
Andreas Oldeweme, MSc
University of Muenster
Universitaetsstrasse 14-16
Muenster, 48143
Germany
Phone: 49 251 83 22991
Email: andreas.oldeweme@wiwi.uni-muenster.de

Abstract

Background: Contact tracing apps are an essential component of an effective COVID-19 testing strategy to counteract the spread of the pandemic and thereby avoid overburdening the health care system. As the adoption rates in several regions are undesirable, governments must increase the acceptance of COVID-19 tracing apps in these times of uncertainty.

Objective: Building on the Uncertainty Reduction Theory (URT), this study aims to investigate how uncertainty reduction measures foster the adoption of COVID-19 tracing apps and how their use affects the perception of different risks.

Methods: Representative survey data were gathered at two measurement points (before and after the app's release) and analyzed by performing covariance-based structural equation modeling (n=1003).

Results: We found that uncertainty reduction measures in the form of the transparency dimensions disclosure and accuracy, as well as social influence and trust in government, foster the adoption process. The use of the COVID-19 tracing app in turn reduced the perceived privacy and performance risks but did not reduce social risks and health-related COVID-19 concerns.

Conclusions: This study contributes to the mass adoption of health care technology and URT research by integrating interactive communication measures and transparency as a multidimensional concept to reduce different types of uncertainty over time. Furthermore, our results help to derive communication strategies to promote the mass adoption of COVID-19 tracing apps, thus detecting infection chains and allowing intelligent COVID-19 testing.

(J Med Internet Res 2021;23(2):e25893) doi:10.2196/25893

KEYWORDS

Uncertainty Reduction Theory; URT; COVID-19; tracing app; mobile health care adoption; DCA-transparency; social influence; initial trust; public health; eHealth; communication; trust; surveillance; monitoring; app; empirical; risk; use

Introduction

Background

At the onset of the COVID-19 pandemic, people, organizations, and governments worldwide were plunged into uncertainty [1], leading to changes in everyday behavior [2,3]. To control the pandemic and protect lives, the authorities have implemented different policies, which range from recommendations (eg, enhanced hand and respiratory hygiene or ventilating rooms) over relatively mild measures (eg, maintaining social distance or mandatory face masks) to far-reaching interventions in civil rights (eg, restrictions in human mobility or lockdowns) [4-6].

Fighting the pandemic effectively is a complex challenge, since limited resources in the health care system and restrictions in everyday life need to be considered simultaneously [7,8]. Consequently, measures to control the pandemic have to be coordinated [9]. Especially when there is no widely available appropriate vaccine, testing, contact tracing, and isolation are considered to be the most essential measures against COVID-19 [9,10]. From a health care management perspective, testing is a key element and provides valuable information regarding the

spread of the virus over time [9]. However, limited testing budgets and test resources such as trained personnel, indicator reagents, polymerase chain reaction devices, or laboratories are the bottlenecks that limit test capacities [11]. Eames and Keeling [12] already showed during the first severe acute respiratory syndrome–related coronavirus pandemic in 2003 that contact tracing, followed by treatment or isolation, is an effective measure in the fight against infectious diseases when testing capacities are limited [13]. Nevertheless, the effectiveness of contact tracing depends primarily on timely and comprehensive collection and processing of data [14,15]. Manual contact tracing only meets these requirements to a limited extent, as it is time and resource intensive, and prone to errors [16]. The resulting urgent need for action toward digital solutions has become evident, as health authorities are collapsing under the burden of manual contact tracing [16].

Contact tracing apps have attracted discussions among politicians, epidemiologists, and the public. These apps aim to systematically identify COVID-19 infection chains and allow a timely and targeted implementation of further measures such as testing and quarantine [14]. Simulation models indicate that digital contact tracing is more efficient compared to manual solutions and has the potential to prevent up to 80% of all transmissions [14,17]. To realize this potential, a majority of the population has to use the same or compatible COVID-19 tracing apps [18]. Achieving mass but voluntary acceptance of the technology is a substantial challenge for several governments [19]. The first positive effects are already expected at a penetration rate of about 20% [15]. However, the penetration rate in several countries as of October 2020 was far below this value, which illustrates the imperative for action to realize the potential of contact tracing apps [20]. Although previous studies have focused on the effectiveness [14,17,19,21] or technical specifications of COVID-19 tracing apps [22], none have examined the factors that affect a rapid and widespread adoption of a COVID-19 tracing app after its release.

Thereby, the use of COVID-19 tracing apps may be related to various uncertainties. These uncertainties can be classified into general health-related COVID-19 concerns and app-specific risks in the form of performance risks and privacy risks that arise because the apps require the processing of sensitive user data [22-24]. In addition, social risks can occur as people might fear social pressure or social exclusion from using or not using the tracing app [25]. According to the Uncertainty Reduction Theory (URT), these uncertainties can be reduced by appropriate means such as transparent communication, social influence, and trust [26-29]. Thereby, the uncertainty reduction means can foster the adoption process of technologies in general [28] and COVID-19 tracing apps in particular. Since COVID-19 tracing apps are mainly released by governments or in cooperation with governmental institutions, trust in the government was examined in addition to the initial trust in a COVID-19 tracing app [24]. For a deeper understanding of the factors at play, the following research questions (RQs) were examined:

- RQ1: How do transparency, social influence, trust in the government, and initial trust in a COVID-19 tracing app affect the adoption process of the app?
- RQ2: What effect does the actual use of COVID-19 tracing apps have on uncertainties in the form of perceived privacy, performance, and social risks, as well as general COVID-19 concerns?

To address the two RQs, we developed a theoretical model based on URT. For testing the model, a representative sample of potential users of a COVID-19 tracing app were surveyed at two different times (1 week before and 4 weeks after the launch of the app) via structured online surveys. Based on this data, we performed covariance-based structural equation modeling (CB-SEM; n=1003). In the following sections, we provide information about COVID-19 tracing apps, explain the theoretical foundation, and derive the hypotheses.

Contact Tracing Apps as a Countermeasure Against COVID-19

This brief review aims to outline the characteristics of automated contact tracing apps for identifying contacts at risk and controlling disease transmission in humans. So far, several countries and some regions have developed and introduced independent COVID-19 tracing apps, which differ in administrative procedure and technical configuration [21]. Two major technical approaches exist: (1) GPS data is used to determine whether individuals, respective to their devices, were located within a geographical proximity for a defined period of time and (2) Bluetooth Low Energy is used to track the concrete proximity and exchange encrypted tokens with other devices in the defined proximity [30,31]. In both cases, data is used to notify people that have been in contact with a person who is infected. The recorded data is either stored on central servers (eg, the tracing app of France) or decentralized locations (eg, the tracing app of Germany) on the particular device [31]. Beside the technical configuration, tracing apps also differ in terms of administrative procedure. Although European tracing apps have been voluntarily used so far, some countries (eg, China) require citizens to install the app [24,32]. Moreover, the source code might be published or withheld by the developers (open source policy). Despite these options, regions need one specific COVID-19 tracing app or at least a suitable interface linking the different apps to achieve a sufficient adoption rate and alert people who are possibly infected [22,33]. A frequently updated overview of the COVID-19 tracing apps used in different regions and their characteristics is provided by MIT Technology Review [20]. Although Trang et al [22] showed that app design influences the likelihood of mass acceptance, there is a lack of evidence to what extent administrative aspects affect the (mass) adoption of COVID-19 tracing apps.

Uncertainty Reduction Theory

URT [26] originally addressed the initial interactions between strangers from a communication science perspective. The core assumption states that individuals face uncertainties in interactions with unknown partners, and individuals attempt to reduce these uncertainties. Berger and Calabrese [26] described uncertainty as a state in which a person is confronted with several alternatives concerning a stranger's behavior. More alternatives make the individuals feel more uncomfortable because the other person's behavior is harder to predict [34]. Although URT was initially developed to explain initial

interactions between individuals, the theory has been applied to other contexts such as recruiting processes [35], computer-mediated communication [36], online commerce [37], or organizational behavior [38]. Hence, URT is not only limited to the interaction of individuals but is also useful in other settings. For instance, Venkatesh et al [28] demonstrated that URT is suitable for explaining the technology-supported communication of individuals and institutions in an e-governance setting. Beyond, URT is also suitable in crisis situations in general and in the current COVID-19 pandemic in particular [29].

The application of URT is appropriate in times of COVID-19 since the situation is marked by various far-reaching uncertainties. Looking at COVID-19 tracing apps, different uncertainties are apparent. First, health care technologies in general often bear uncertainties concerning data privacy [39]. These uncertainties are also identified in cases of COVID-19 tracing apps, as they require the processing of sensitive personal data [22,30]. Individuals fear that their privacy will be violated and cause undesirable outcomes such as governmental surveillance [30,40]. Moreover, people are concerned that personal data is used to impose quarantine or restrict access to public places for people who do not use a COVID-19 tracing app [24]. Second, uncertainties about the true performance and functionality of tracing apps are apparent [23]. Using a mobile app to contain a pandemic is new to individuals in most countries. Hence, they cannot draw on past experiences and might question its utility (eg, false alerts or only few people using the app). Third, social risks are recognizable as people might fear social pressure or social exclusion from using or not using the tracing app [25]. Beside tracing app–related uncertainties, general health-related COVID-19 concerns arise from the pandemic itself. The main aspects of the four uncertainties considered are summarized in Table 1. In addition, the four described uncertainties are further reinforced by unverified information and fake news [41-43].

Table 1. Summary of relevant uncertainties in the context of COVID-19 tracing apps.

Relevant uncertainties	Related to tracing apps	Description
Privacy risks	Yes	Individuals are uncertain about data security (ie, possible data leaks or misuse by third parties). Hence, tracing apps are perceived as risky because they bear the potential loss of control over personal data [22,44].
Performance risks	Yes	Individuals are concerned that the product may not work and perform as it was designed and advertised. As a result, people are uncertain whether enough people will use apps for contact tracing and whether the technology will work as expected [23].
Social risks	Yes	Individuals might fear potential loss of status in one's social group for using or not using the app. In addition, forced quarantine might lead to social isolation [25].
COVID-19 concerns	No	Individuals worry about negative impacts arising from the COVID-19 pandemic. Fear and anxiety about a new disease, for their own health and their relatives, can be overwhelming [45].

Uncertainty Reduction Measures

According to URT, individuals reduce uncertainties by passive (observing), active (target-orientated search), and interactive (interaction with the stranger) information-seeking approaches [26,27]. We discuss transparency, social influence, and (initial) trust, as these factors facilitate individual's information-seeking strategies [28,46-48].

Notably for passive and active strategies, individuals rely on accessible and valuable information [46]. To enable people to reduce uncertainty through observation or targeted research, information must be available. If no information is obtainable, people cannot reduce uncertainties through observation or targeted research. Therefore, transparency is examined as an enabler for passive and active information-seeking strategies. We defined transparency as "the perceived quality of intentionally shared information from a sender" [49]. Drawing on recent transparency research, transparency is best understood as a multidimensional construct consisting of disclosure, clarity, and accuracy of information [49-51]. In the context of this study, disclosure is the perception that sufficient relevant information about a COVID-19 tracing app is timely and accessible. Similarly, clarity is the perception that the received information about a COVID-19 tracing app is comprehensible and lucid. For instance, the disclosure of a huge amount of information cannot be considered transparent if the information is not understandable for individuals (eg, because the information is cryptic and only consists of the technical code of the COVID-19 tracing app). This information would hinder an individual's ability to effectively perform active and passive information seeking. Lastly, accuracy is the perception that the information about a COVID-19 tracing app is correct [49]. The apparent incorrectness of information would not lower uncertainty but might lead to concerns about hidden governmental intentions. Notably in the context of a pandemic, each transparency dimension contributes to the reduction of uncertainty, as individuals rely on sufficient, relevant, timely, clear, and accurate information to observe the unknown technology and to actively search for information [29].

Furthermore, interactive information-seeking approaches have been shown to be more efficient than passive or active strategies in reducing uncertainty [46]. As it is not possible to interact with COVID-19 tracing apps before they are released or to directly communicate with the people responsible for the app, people seek alternatives for interactive information gathering. Therefore, individuals may communicate with their peers who

are also affected by the decision whether to use the app or not. This, in turn, has to be regarded as another active information-seeking approach rather than an interactional strategy. Although communication with the social environment is interactive, the social environment is not the publisher of the app, and therefore, referring to URT, social influence is an active information-seeking approach. Social influence is expected to reduce people's uncertainty about COVID-19 tracing apps, and it is defined "as the degree to which an individual perceives that important others believe he or she should use the new system" [52]. By knowing the preferences of their social environment, individuals' attitudes toward using the app might be affected.

Lastly, trust is shown to reduce uncertainties and risks in different settings [53,54], and it is defined as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behaviors of another" [55]. We distinguish between initial trust in COVID-19 tracing apps and individuals' trust in their government. Several positive links to uncertainty reduction exist for initial trust in new technologies. For example, in e-commerce, trust lowers customers' uncertainties about vendor behavior [56,57]. Additionally, initial trust reduces citizens' uncertainties in the wake of e-governance [25]. However, initial trust might change in the actual use of the app and become strengthened or weakened according to the specific experiences encountered [58]. In addition to initial trust, individuals' trust in their government is another means to reduce uncertainties. As most COVID-19 tracing apps are published by governments, trust in the administration might reduce fears related to app use [59]. People's trust in the government is expected to be relatively stable and not fundamentally changeable in the short term [60,61].

Transparency and Initial Trust

Based on the transparency and trust literature, it is widely believed that transparency perceptions are positively related to trust [62]. This is shown by Schnackenberg et al [63] who explored the positive role of employees' transparency perceptions (disclosure, clarity, accuracy) in the context of employees' trust in their manager in organizational settings. Rawlins [51] also showed a positive link between transparency and employee trust, and highlighted the mutual relation between transparency and trust. Regarding the consequences of corporate scandals, transparency can be used as a strategic tool to restore stakeholder trust in firms [64]. In financial markets, transparency is shown to influence citizens' trust in central banks [65]. In the case of COVID-19 tracing apps, a certain degree of transparency must be achieved for people to trust the app and use the technology [24]. The formation of peoples' initial trust in COVID-19 tracing apps relies on the quality of available information as long as there are no prior interactions between citizens and the app [53,57]. Fulfilling certain information needs (eg, by providing sufficient clear and accurate information) enables people to initially trust a COVID-19 tracing app.

- Hypothesis (H)1: (a) Disclosure, (b) clarity, and (c) accuracy are positively related to individuals' initial trust in a COVID-19 tracing app.

Trust in the Government and Initial Trust

Trust transfer theory states that individuals' trust in a specific area can influence initial trust in other domains that are believed to have certain links to the known and trusted domain [66]. For instance, Lu et al [67] demonstrated that customers' trust in internet payment in general influences trust in mobile payment services. As the majority of COVID-19 tracing apps are published by government institutions, trust in the government might affect initial trust in a COVID-19 tracing app. Peoples' trust in the government is defined as the "perceptions regarding the integrity and ability of the agency providing [a] service" [53]. When people believe that the government is generally acting in citizens' best interest and when citizens perceive the government agencies as capable to appropriately offer services, the initial trust in a COVID-19 tracing app is strengthened [53]. Recent studies on COVID-19 tracing apps noted that trust in the government influences peoples' attitude toward the specific app [59,68]. In addition, a main reason for general negative attitudes against COVID-19 tracing apps is a lack of trust in the government [68]. Therefore, based on trust transfer theory, trust in the government fosters peoples' initial trust in a COVID-19 tracing app [28].

- H2: Trust in the government is positively related to peoples' initial trust in a COVID-19 tracing app.

Social Influence, Initial Trust, and Intention to Use

As previously stated, social influence can also aid the understanding of uncertainty reduction, as it might function as a substitute for interaction with the unknown and not yet available technology. Therefore, social interaction serves as an active means to gather information. The presumed reactions of the social environment will influence an individual's attitude and behavior in a technology adoption context [52]. In terms of URT, people access their social environment as an active information-seeking means by interacting with their peers to exclude possible consequences of using or not using the specific technology. In this sense, social interaction, just like transparency, is a means of obtaining information and excluding alternatives and, hence, serves to reduce uncertainties. Li et al [69] showed that social influence is an important factor for the formation of initial trust and is therefore contributing to the exclusion of expectable negative outcomes such as perceived risks. Against this background, we argue that initial trust in a COVID-19 tracing app is not only influenced by transparency and trust in government, but is also affected by social influence.

- H3a: Social influence is positively related to individuals' initial trust in a COVID-19 tracing app.

In addition, it is well known that social influence is an important antecedent of intention to use new technologies [70-72]. Being part of social groups (eg, family or colleagues) creates pressure on individual behavior, as people try to behave in accordance to established standards [73]. In health care settings, it has been shown that social influence is, for example, leading to smoking cessation [74] or supporting to maintain a diet [73,75]. Besides positive effects, the peer group might also foster negative behaviors such as drug abuse [76]. Therefore, social influence is a major factor to consider in the adoption process of health

care technologies. This is particularly reinforced in the case of preventive behaviors like using tracing apps whose positive effect is not directly evident [73]. Therefore, we expect that social influence is not only influencing one's initial trust in COVID-19 tracing apps but also impacts one's intention to use the technology.

- H3b: Social influence is positively related to individuals' intention to use a COVID-19 tracing app.

Initial Trust and Intention to Use

In the technology acceptance literature, trust has been shown to be positively correlated with the intention to use technology [69]. For instance, Nicolaou and McKnight [77] demonstrated that individuals' trusting beliefs increase the intention to engage in interorganizational information exchange. Furthermore, (initial) trust is identified to be an antecedent of citizens' intention to use e-governance services [28,53]. Parker et al [24] argued that the successful launch of mobile apps to fight the COVID-19 pandemic in democratic countries relies on the ability to ensure peoples' trust in the technology. Based on URT, we argue that initial trust is a means to exclude potential negative behavior of the technology provider. Citizens who trust a COVID-19 tracing app estimate the probability of deceitful intentions as low.

- H4: Initial trust in a COVID-19 tracing app is positively related to individuals' intention to use it.

Intention to Use and Actual Use

According to the theory of planned behavior [78], established technology acceptance theories (eg, technology acceptance model [TAM] or unified theory of acceptance and use of technology [UTAUT]) [52,79], and the application of URT in the technology context [28], explains that individuals' actual use of new technology is influenced by individuals' intention to use the technology [80]. This relationship is also expected to be applicable in the context of COVID-19 tracing apps.

- H5: Intention to use is positively related to individuals' actual use of a COVID-19 tracing app.

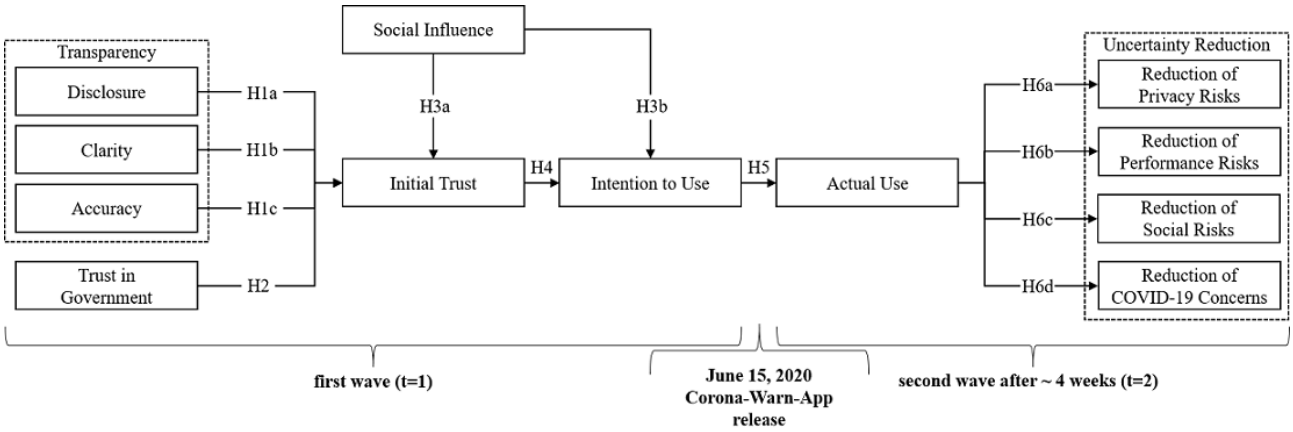
Actual Use and Uncertainty Reduction

Referring to URT, uncertainties concerning data privacy, app performance, social consequences, and general COVID-19 concerns are decreased by the aforementioned means during the adoption process. The actual use of a COVID-19 tracing app is the only available interactive information-seeking possibility for individuals. Therefore, it is effective as it involves a direct interaction with the unknown technology [27]. This interaction enables individuals to discard uncertainties such as performance uncertainties (eg, functionality and handling) of COVID-19 tracing apps [29]. However this mean can obviously only be used after the app has been released. The investigation of the relationship between actual use and the reduction of different forms of uncertainty further addresses the applicability of URT for technology acceptance in uncertain environments.

- H6: The actual use of a COVID-19 tracing app is positively related to the reduction of (a) privacy risks, (b) performance risks, (c) social risks, and (d) COVID-19 concerns.

The proposed research model is summarized in Figure 1.

Figure 1. Proposed research model. H: hypothesis.



Methods

Data and Procedure

To investigate the adoption process of COVID-19 tracing apps and test the theoretical model, data on the German "Corona-Warn-App" were collected via the panel platform respondi in two waves using structured online surveys. Members of this panel voluntarily agree to receive invitations to scientific surveys and may unsubscribe or delete their personal information at any time. Participants were assigned a randomly generated identifier, allowing us to match the results of both surveys. Nevertheless, since a third party (respondi) collected the data for both waves, we did not have direct contact with participants or access to identifying participant information. In addition, the survey did not collect any personal identifying information. Consequently, we were able to guarantee the anonymity and privacy of the participants at all times and acted in accordance with the ethical principles of the German Research Foundation.

The data collection of the first wave (t1) lasted 1 week and was completed 1 day before the app was released on June 15, 2020. Prior to the survey, the participants received the official information from the Federal German Government about the app [81]. For detailed information, see Multimedia Appendix

1 [81]. The app was developed on Android and iOS platforms by Robert Koch-Institut in conjunction with the private companies SAP, Deutsche Telekom, and other partners on behalf of the German government [81]. For the administrative conditions, the source code is publicly available on GitHub [82,83]. Furthermore, no registration is required to run the app, and the use is voluntary. The app uses Bluetooth Low Energy technology to record contact between people. Considering data privacy and security concerns, the devices exchange temporary encrypted random codes (Bluetooth ID) with each other [84]. The use of random codes prevents conclusions from being drawn about individual users or their specific locations. These codes are stored decentralized on the mobile devices, and the tracing data collected are automatically deleted after 14 days [85].

Subsequently, the participants received a quantitative questionnaire concerning their evaluations and perceptions of the app as well as demographical information. In the second survey wave (t2), the same participants were surveyed again 4 weeks after the app was released. The actual use of the app was queried in addition to their evaluations and perceptions.

To maintain data quality and ensure scale validity, we included three attention checks and screened out participants that failed these tests. Finally, 1373 individuals completed the first survey, and 1050 participated in the second wave, yielding a completion rate of 76.47% (1050/1373). Owing to four "knock-out" criteria regarding low response times, missing data, suspicious response patterns, and outliers, 47 participants were excluded from the analysis to ensure data quality [86]. A final sample size of 1003 was then obtained. This sample is representative of Germany in terms of gender, age, and education (see Table 2).

Table 2. Representativeness description.

Characteristic	Germany, %	Sample, n (%)
Gender		
Male	50	490 (48.85)
Female	50	512 (51.05)
Divers	0	1 (0.10)
Age (years)		
14-29	22	190 (18.94)
30-39	16	138 (13.76)
40-49	16	168 (16.75)
50-59	20	214 (21.34)
≥60	25	291 (29.01)
Education		
Low	36	368 (36.69)
Middle	31	315 (31.41)
High	33	320 (31.90)
Adoption rate of Corona-Warn-App (August 2020)[a]	27	379 (37.80)

[a]Calculation of adoption rate in Germany: 16.6 million downloads / approximately 62 million smartphone users (source: Statistisches Bundesamt [87,88], and Robert Koch-Institut [83]).

Measures

To test the proposed research model, we used established scales that have been validated in previous studies. Except for demographics, use behavior (binary), and control variables (gender, age, education), the participants rated all items on 5-point Likert scales. Intention to use was measured by a 3-item scale [52]. For initial trust in the Corona-Warn-App, we built on a 5-item scale developed by Koufaris and Hampton-Sosa [58]. Trust in government was examined through a 4-item scale adapted from Bélanger and Carter [53]. Transparency (reflecting individuals' perceptions of information quality) was adopted from Schnackenberg et al [63], and each dimension was based on 4 items. Privacy risks was measured on a 5-item scale developed by Rauschnabel et al [44], and a 5-item measure was adopted from Featherman and Pavlou [89] to assess performance and social risks. Finally, general COVID-19 concerns were measured through a 6-item scale by Conway et al [45].

We calculated differences for the four dimensions of risk perceptions between the two survey waves to measure the change in the perceived risk assessments. The differences were calculated using the following formula: difference variables = risk perception$_{(t1)}$ – risk perception$_{(t2)}$. The means, SDs, and correlations for all constructs are reported in Table 3. Age, gender, and education were used as controls.

Table 3. Mean, SD, and correlations.

Variables	Mean (SD)	1	2	3	4	5	6	7	8	9	10	11
1. Disclosure	3.158 (0.909)											
Correlation		__[a]										
P value		—										
2. Clarity	3.647 (0.834)											
Correlation		0.645	—									
P value		<.001	—									
3. Accuracy	3.566 (0.898)											
Correlation		0.587	0.705	—								
P value		<.001	<.001	—								
4. Social influence	2.841 (1.12)											
Correlation		0.440	0.427	0.585	—							
P value		<.001	<.001	<.001	—							
5. Trust in government	3.13 (0.979)											
Correlation		0.332	0.334	0.505	0.454	—						
P value		<.001	<.001	<.001	<.001	—						
6. Initial trust	3.147 (1.081)											
Correlation		0.551	0.545	0.742	0.71	0.59	—					
P value		<.001	<.001	<.001	<.001	<.001	—					
7. Intention to use	3.022 (1.444)											
Correlation		0.434	0.429	0.619	0.685	0.466	0.803	—				
P value		<.001	<.001	<.001	<.001	<.001	<.001	—				
8. Actual use	1.378 (0.485)											
Correlation		0.288	0.307	0.377	0.419	0.355	0.512	0.595	—			
P value		<.001	<.001	<.001	<.001	<.001	<.001	<.001	—			
9. Privacy risk	—											
Correlation		–0.074	0.037	–0.033	–0.100	–0.047	–0.106	–0.070	0.224	—		
P value		.02	.24	.30	.002	.14	<.001	.03	<.001	—		
10. Performance risk	—											
Correlation		–0.063	–0.001	–0.055	–0.148	–0.008	–0.109	–0.091	0.169	0.595	—	
P value		.045	.97	.08	<.001	.80	<.001	.004	<.001	<.001	—	
11. Social risk	—											
Correlation		0.031	–0.044	–0.020	0.064	–0.023	–0.001	0.025	0.008	0.007	0.086	—
P value		.32	.16	.54	.04	.47	.98	.43	.81	.83	.006	—
12. COVID-19 concerns	—											
Correlation		0.070	0.081	0.038	0.057	0.042	0.04	0.033	0.011	–0.023	0.027	0.041
P value		.03	.01	.23	.07	.19	.21	.30	.72	.48	.39	.19

[a]Not applicable.

Before conducting the structural equation modeling (SEM) analysis, we tested the reliability and validity of the measurement model. One item displayed poor factor loadings and was dropped (TR_5). All other factor loadings exceeded the threshold of 0.6. Internal consistency and composite reliability were assumed, as the Cronbach alpha met the quality criteria of >.7, and the average variance extracted exceeded 0.5 [90,91]. Composite reliability of all items exceeded the cut-off value of 0.6 [92]. The final questionnaire with all constructs, related survey items, their sources, and the aforementioned indexes is presented in Multimedia Appendix 2 [44,45,52,53,58,63,89,93].

Data Analysis

We used the R-based JASP software (University of Amsterdam) environment to evaluate our proposed research model [94] and the lavaan code to conduct CB-SEM [95] analysis. Before performing the SEM analyses, we tested the fit, reliability, and validity of the applied model. The comparative fit index (>0.95), Tucker-Lewis index (>0.95), root mean square error of approximation (<0.08), and standardized root mean square residuals (<0.08) complied with the conventional cut-off criteria [96,97]. Based on Kline [98], the χ^2 / df ratios indicated a sufficient model fit across models (<3). Common method bias was not a problem, as the Harman single factor test indicated that only a variance of 27.6% were explained by a single factor consisting of all model items [99]. In summary, all fit indexes revealed a very good overall model fit (see Table 4), with all indicators reaching their respective thresholds.

Table 4. Covariance-based structural equation modelling results.

Items	β (SE)[a]	*P* value	Assessment of hypotheses	Index values
Hypotheses				N/A[b]
H[c]1a	.140 (.030)	<.001	Supported	
H1b	–.028 (.041)	.45	Rejected	
H1c	.375 (.035)	<.001	Supported	
H2	.201 (.022)	<.001	Supported	
H3a	.377 (.025)	<.001	Supported	
H3b	.207 (.033)	<.001	Supported	
H4	.670 (.035)	<.001	Supported	
H5	.599 (.013)	<.001	Supported	
H6a	.222 (.048)	<.001	Supported	
H6b	.169 (.049)	<.001	Supported	
H6c	.005 (.064)	.88	Rejected	
H6d	.012 (.031)	.72	Rejected	
Controls			N/A	N/A
Age → PRPP[d]	–.090 (.002)	.005		
Age → PR[e]	–.062 (.002)	.06		
Age → SR[f]	–.004 (.002)	.91		
Age → CC[g]	–.064 (.000)	.06		
Gender → PRPP	–.001 (.047)	.96		
Gender → PR	.000 (.048)	.99		
Gender → SR	.015 (.062)	.63		
Gender → CC	–.017 (.030)	.59		
Education → PRPP	–.006 (.020)	.85		
Education → PR	.000 (.021)	.99		
Education → SR	.020 (.027)	.54		
Education → CC	–.007 (.013)	.84		
Indexes	N/A	N/A	N/A	
Comparative fit index				0.975
Tucker-Lewis index				0.972
RSMEA[h]				0.040
SRMR[i]				0.057
Chi-square (*df*)				1270.187 (491)
Chi-square / *df*				2.587

[a]Standardized path coefficients; standard error of the estimators in parentheses.

[b]N/A: not applicable.

[c]H: hypothesis.

[d]PRPP: reduction of privacy risks.

[e]PR: reduction on performance risks.

[f]SR: reduction of social risks.

[g]CC: reduction of COVID-19 concerns.

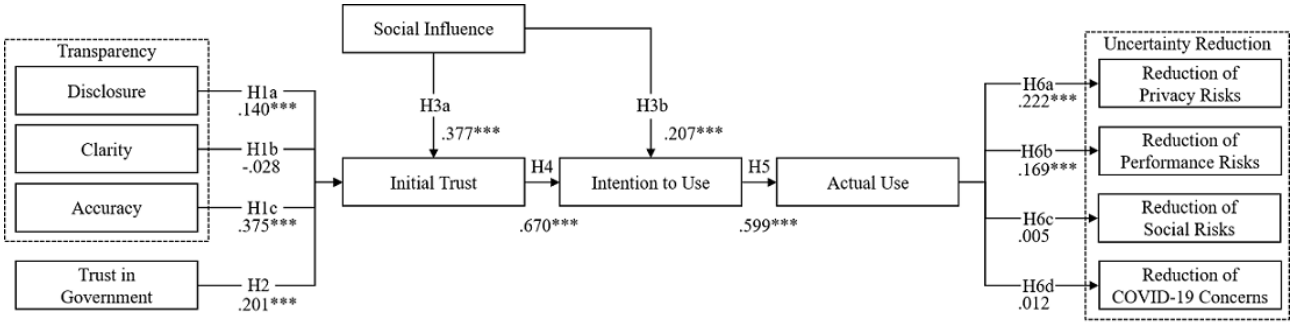
[h]RSMEA: root mean square error of approximation.

[i]SRMR: standardized root mean square residuals.

Results

The standardized path coefficients, significance levels, and fit indexes are summarized in Table 4. As illustrated in Figure 2, information disclosure and accuracy are positively related to initial trust, supporting H1a (β=.140; *P*<.001) and H1c (β=.375; *P*<.001). In contrast, H1b was rejected (β=–.028; *P*=.45), as information clarity shows no relation to initial trust. H2 was supported (β=.201; *P*<.001) as trust in governance and initial trust were positively related. Furthermore, there was support for H3a (β=.377; *P*<.001) and H3b (β=.207; *P*<.001), as the results showed a positive relation between social influence toward initial trust and intention to use. The observed relationship between initial trust and intention to use was positive, supporting H4 (β=.670; *P*<.001). We also found support for H5 (β=.599; *P*<.001), as intention to use was positively related to the actual use of a COVID-19 tracing app. Finally, we found a positive relationship between actual use and privacy and performance risks, thus supporting H6a (β=.222; *P*<.001) and H6b (β=.169; *P*<.001). In contrast, H6c (β=.005; *P*=.88) and H6d (β=.012; *P*=.72) were rejected as actual use was not related to social risks or COVID-19 concerns, respectively. The control variables gender and education were not related to the reduction of the four dimensions of uncertainty reduction, while age was negatively related to privacy risk reduction.

Figure 2. Results of structural equation model. H: hypothesis. ***P<.001.



In addition to the hypothesized direct relationships between social influence and the intention to use the COVID-19 tracing app, we conducted a post hoc analysis to investigate the potential indirect effects of social influence on the intention to use the COVID-19 tracing app mediated by initial trust. The mediation effects were examined with the help of the procedure according to Baron and Kenny [100] and are depicted in Table 5. We found evidence that the indirect effect was significant (β=.253; *P*<.001). As the direct effect of social influence toward the intention to use has been shown to be significant before (H5), we postulated that the relationship of social influence and intention to use is partially mediated by initial trust.

Table 5. Mediating effect.

Effect	β (SE)[a]	*P* value	Mediation
Indirect effect: social influence → initial trust → intention to use	.253 (.022)	<.001	partial mediation
Total effect: social influence → intention to use	.460 (.029)	<.001	–

[a]Standardized path coefficients; standard error of the estimators in parentheses.

Discussion

General Discussion

In this study, we investigated how uncertainty reduction measures can foster the adoption of COVID-19 tracing apps and, consequently, the reduction of uncertainty perception. In this section, we discuss the antecedents of initial trust, intention, and actual use of the app, as well as the reduction of specific uncertainties. Based on URT, transparency and social influence are antecedents of initial trust. In terms of transparency, we found that initial trust in COVID-19 tracing apps is positively influenced by the disclosure and accuracy of information. However, accuracy had a considerably higher effect on initial trust than the disclosure dimension. This shows that, although it is important to receive sufficient information, the perceived validity of the information is crucial. Unexpectedly, we found no effect between information clarity and initial trust in the Corona-Warn-App. This may be due to the peculiarities of the COVID-19 pandemic, as people likely became used to constantly encountering new complex information and thus accepted a lower level of information lucidity. Despite the missing effect between clarity and initial trust, our findings are consistent with the existing transparency-trust literature [51,62,63].

As proposed, social influence positively affects individuals' initial trust. The integration of social influence in the URT context reveals that social influence serves as an active information-seeking strategy, thus meeting the demand of Venkatesh et al [28] to integrate UTAUT variables into URT. Especially in situations where direct interaction with the unknown technology is not possible, the communication with peers becomes important. In addition, we identified a positive relation between social influence and intention to use. This is in line with technology acceptance and health care literature [52,70,73]. In addition, we were able to show that initial trust partially mediates the relationship between social influence and intention to use COVID-19 tracing apps. Furthermore, we examined the effect of trust in the government on initial trust

in COVID-19 tracing apps and found a positive relationship between these concepts. This is consistent with the trust transfer theory and current studies on COVID-19 tracing apps [59,68]. It is important to note that trust in government has a smaller effect on individuals' initial trust compared to transparency and social influence. Therefore, people who are critical of the government can still develop initial trust in the app through other short-term influenceable means such as transparent communication.

Additionally, we observed a positive relation between initial trust and intention to use. This result is consistent with URT [27,28] and confirms the common understanding of trust in the context of technology acceptance (for a meta-analysis, see Wu et al [101]). As expected, people who have a high intention to use a COVID-19 tracing app are more likely to use it. Nevertheless, our results also revealed, as most studies have, that an intention-behavior gap exists [102].

Considering uncertainty reduction specifically, we found that the actual app use increases COVID-19 tracing app–related uncertainty reduction. Individuals' uncertainty reduction of perceived privacy and performance risks was significantly increased by using the app. Thus, we found support for Trang et al [22], who stated that data privacy and app performance (benefits) need to be considered in the development of tracing apps. In addition, our results did not indicate a reduction of social risks nor a reduction of general COVID-19 concerns. As COVID-19 concerns span broader health-related fears, they cannot be solely linked to the functionality of the app or interaction with it. Tracing apps do not provide direct protection but are mainly intended to identify infection chains to implement further appropriate actions such as intelligent testing and quarantine [9]. This explains why the use of a COVID-19 tracing app has no impact on the reduction of these general health-related fears. Further, it indicates that people using tracing apps are not getting more reckless but still recognize the virus's threat. Regarding social risks, the use and nonuse of the app is less visible to nearby people than wearing a face mask or complying with social distance regulations. Therefore, individuals' actual use behaviors might be unrelated to social consequences as long as the use of such an app is not mandatory, for example, to use public transportation or enter restaurants or other places. For the controls, we found that age was negatively correlated to the reduction of privacy risks. This effect is rather small and in line with research emphasizing that privacy concerns are more pronounced and stable among older people than among younger individuals [103].

Theoretical Implications

Our study design and findings contribute to the literature in several ways. First, we demonstrated with our study design how mass adoption problems can be investigated over time in the health care management context using the example of a COVID-19 tracing app. By applying URT, we contributed to its empirical validation in general and introduced it to the field of health care management. The application is particularly valuable in the health care context, as this area is characterized by uncertainties that may lead to serious and far-reaching consequences, as is apparent in times of the COVID-19 pandemic [104]. Second, it was shown that interactive information-seeking strategies, such as app use, are appropriate for reducing related uncertainties (eg, privacy and performance risks). By collecting the data in two measurement periods (before and after the release of the app) and calculating difference variables to quantify the uncertainty reduction, we validated the impact of the use of a technology on uncertainty reduction. The use of specific uncertainty reductions as outcome variables is theoretically stronger for URT than the use of outcome variables such as satisfaction proposed by Venkatesh et al [28]. Third, further theoretical contributions were made by integrating recent transparency research [49] into URT. Thereby, our results highlighted the importance of considering transparency as a multidimensional construct [49]. Transparency perceptions are essential as they form the basis for active and passive information-seeking strategies. By using the recent DCA-transparency scale [63], we further elaborated on the role of transparency (ie, information quality) in URT as proposed by Venkatesh et al [28]. Finally, it was shown that trust transfer theory holds true in the investigated setting. Although trust in the government is not a major antecedent for initial trust in COVID-19 tracing apps, individuals' trust in the government should still be considered in governmental technology publishing.

Implications for Practice

The adoption rates of voluntary COVID-19 tracing apps differ largely among countries and are mostly below the critical thresholds, which hinders their effectiveness [14,20]. To improve acceptance, governments can adopt the following implications in their communication strategies. First, governments that introduce a voluntary COVID-19 tracing app (or other technologies) should engage in a transparent communication process. A supply of sufficient information, which must be perceived as accurate, is thus required. However, transparent communication only works if the service itself exceeds certain standards such as data privacy and security [22]. Second, interactive information-seeking strategies of individuals must be managed. These strategies (eg, app use) are shown to be efficient in terms of uncertainty reduction. Hence, governments should provide appropriate formats to enable interactive information seeking before release. Such formats can be demo versions, realistic previews, question and answer sessions, or even hackathons. Finally, our findings are extendable to other technologies and settings. For example, if there are other digital trends in the health care system (eg, digital health record or video doctor), our results can be applied to achieve (voluntary) technology (mass) acceptance. Whenever governments or organizations develop and publish new services (eg, disaster alarm app), other uncertainties such as financial risk, time risk, or psychological risk may arise and should be considered. The conscious management of the (transparent) publication process can promote a successful launch of a technology. By understanding the multidimensional nature of perceived information quality, both organizations and governments can reflect and develop their own technology implementation strategy. Hence, many of the implications outlined here may also be relevant to future pandemics and public health crises.

Limitations and Future Research

Although the results of this study provide important insights, the study has some limitations. As the results are based on data related to the German COVID-19 tracing app, the generalizability of our findings for other regions may be restrained due to cultural differences. Thus, future research should expand this study by including other countries. Further, actual app use was self-reported by the participants and might be untrue in some cases. However, the app adoption rate in our sample was comparable with the adoption rate in the German population during the second survey wave (see Table 2). To advance URT, researchers can examine the communication channels that are most suitable to ensure transparency and reduce different uncertainties. After some studies have dealt with the design [105], the technical configuration [22], and the ethical guidelines [106], we studied the requirements for adequate app implementation and communication. Therefore, future research should investigate means to ensure mid- and long-term app acceptance and use.

For most of the population, the Corona-Warn-App was a new concept at the time of its release. Since then, the app and its functionality have become relatively well known and widespread. For this reason, follow-up research should investigate the role of descriptive norms (ie, how others actually behave) besides subjective norms, which we have investigated in the form of social influence (ie, how important others think one should behave), for the adoption process [107].

Moreover, the data underlying this study originated a few days (t1) before and 4 weeks (t2) after the launch of the COVID-19 tracing app in Germany and, thus, between the first and second waves of infection. In the meantime, various measures against the pandemic have been implemented, and more information about the virus, its spread, and mortality are available. These insights should be considered in follow-up studies. For example, the distribution and adoption of new SARS-CoV-2 vaccines represent a milestone in the fight against the pandemic. Therefore, follow-up studies should examine whether these insights influence the use of the COVID-19 tracing app and uncertainty perceptions.

Conclusion

A key strategy in fighting the COVID-19 pandemic is the testing and subsequent isolation of individuals who are potentially infected. The automatic contact tracing via mobile apps offers an important contribution to the decision of which people need to be tested with regard to limited testing capacities. Our study offers original insights on the factors driving the mass acceptance of COVID-19 tracing apps to identify infection chains and control the pandemic. Building on URT and through a longitudinal empirical study on the adoption process, we investigated how uncertainty reduction measures affect the adoption of COVID-19 tracing apps and how their use affects the perception of different risks. We analyzed representative data through CB-SEM. The results revealed that the transparency dimensions of disclosure and accuracy, as well as social influence, trust in government, and initial trust positively affect the adaptation process, whereas no effect was observed for the transparency dimension clarity. Further, we showed that the actual use of COVID-19 tracing apps reduces the perceived uncertainty regarding performance and privacy risks, but no effect on the reduction of social risks and COVID-19 concerns was identified. Finally, we derived theoretical and practical implications concerning the communication strategy of contact tracing apps in particular and for health care technologies in general.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Background information for participants.
[DOCX File , 14 KB - jmir_v23i2e25893_app1.docx]

Multimedia Appendix 2
Items of the survey and their sources.
[DOCX File , 20 KB - jmir_v23i2e25893_app2.docx]

References

1. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
2. West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. Nat Hum Behav 2020 May;4(5):451-459. [doi: 10.1038/s41562-020-0887-9] [Medline: 32377018]
3. Harper CA, Satchell LP, Fido D, Latzman RD. Functional fear predicts public health compliance in the COVID-19 pandemic. Int J Ment Health Addict 2020 Apr 27:1-14 [FREE Full text] [doi: 10.1007/s11469-020-00281-5] [Medline: 32346359]
4. Altmann DM, Douek DC, Boyton RJ. What policy makers need to know about COVID-19 protective immunity. Lancet 2020 May 16;395(10236):1527-1529 [FREE Full text] [doi: 10.1016/S0140-6736(20)30985-5] [Medline: 32353328]

5. Rocklöv J, Sjödin H, Wilder-Smith A. COVID-19 outbreak on the Diamond Princess cruise ship: estimating the epidemic potential and effectiveness of public health countermeasures. J Travel Med 2020 May 18;27(3):taaa030 [FREE Full text] [doi: 10.1093/jtm/taaa030] [Medline: 32109273]
6. Chiu N, Chi H, Tai Y, Peng C, Tseng C, Chen C, et al. Impact of wearing masks, hand hygiene, and social distancing on influenza, enterovirus, and all-cause pneumonia during the coronavirus pandemic: retrospective national epidemiological surveillance study. J Med Internet Res 2020 Aug 20;22(8):e21257 [FREE Full text] [doi: 10.2196/21257] [Medline: 32750008]
7. Usher K, Bhullar N, Jackson D. Life in the pandemic: social isolation and mental health. J Clin Nurs 2020 Aug;29(15-16):2756-2757. [doi: 10.1111/jocn.15290] [Medline: 32250493]
8. Wood RM, McWilliams CJ, Thomas MJ, Bourdeaux CP, Vasilakis C. COVID-19 scenario modelling for the mitigation of capacity-dependent deaths in intensive care. Health Care Manag Sci 2020 Sep;23(3):315-324 [FREE Full text] [doi: 10.1007/s10729-020-09511-7] [Medline: 32642878]
9. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 2020 Jun;26(6):855-860 [FREE Full text] [doi: 10.1038/s41591-020-0883-7] [Medline: 32322102]
10. Salathé M, Althaus CL, Neher R, Stringhini S, Hodcroft E, Fellay J, et al. COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. Swiss Med Wkly 2020 Mar 09;150:w20225. [doi: 10.4414/smw.2020.20225] [Medline: 32191813]
11. Esbin MN, Whitney ON, Chong S, Maurer A, Darzacq X, Tjian R. Overcoming the bottleneck to widespread testing: a rapid review of nucleic acid testing approaches for COVID-19 detection. RNA 2020 Jul;26(7):771-783 [FREE Full text] [doi: 10.1261/rna.076232.120] [Medline: 32358057]
12. Eames KTD, Keeling MJ. Contact tracing and disease control. Proc Biol Sci 2003 Dec 22;270(1533):2565-2571 [FREE Full text] [doi: 10.1098/rspb.2003.2554] [Medline: 14728778]
13. Kaplan EH. Containing 2019-nCoV (Wuhan) coronavirus. Health Care Manag Sci 2020 Sep;23(3):311-314 [FREE Full text] [doi: 10.1007/s10729-020-09504-6] [Medline: 32146554]
14. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
15. Armbruster B, Brandeau ML. Contact tracing to control infectious disease: when enough is enough. Health Care Manag Sci 2007 Dec;10(4):341-355 [FREE Full text] [doi: 10.1007/s10729-007-9027-6] [Medline: 18074967]
16. Keeling MJ, Hollingsworth TD, Read JM. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). J Epidemiol Community Health 2020 Oct;74(10):861-866 [FREE Full text] [doi: 10.1136/jech-2020-214051] [Medline: 32576605]
17. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. Lancet Public Health 2020 Aug;5(8):e452-e459 [FREE Full text] [doi: 10.1016/S2468-2667(20)30157-2] [Medline: 32682487]
18. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. JMIR mHealth uHealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
19. Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-peer contact tracing: development of a privacy-preserving smartphone app. JMIR mHealth uHealth 2020 Apr 07;8(4):e18936 [FREE Full text] [doi: 10.2196/18936] [Medline: 32240973]
20. O'Neill PH, Ryan-Mosley T, Johnson B. A flood of coronavirus apps are tracking us. Now it's time to keep track of them. MIT Technology Review. 2020 May 07. URL: https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/ [accessed 2020-08-05]
21. Urbaczewski A, Lee YJ. Information technology and the pandemic: a preliminary multinational analysis of the impact of mobile tracking technology on the COVID-19 contagion control. Eur J Inf Syst 2020 Aug 10;29(4):405-414. [doi: 10.1080/0960085x.2020.1802358]
22. Trang S, Trenz M, Weiger WH, Tarafdar M, Cheung CM. One app to trace them all? Examining app specifications for mass acceptance of contact-tracing apps. Eur J Inf Syst 2020 Jul 27;29(4):415-428. [doi: 10.1080/0960085x.2020.1784046]
23. Fenton NE, Mclachlan S, Lucas P, Kudakwashe D, Hitman GA, Osman M, et al. A Bayesian network model for personalised COVID19 risk assessment and contact tracing. medRxiv Preprint posted online January 5, 2021. [FREE Full text]
24. Parker MJ, Fraser C, Abeler-Dörner L, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics 2020 Jul;46(7):427-431 [FREE Full text] [doi: 10.1136/medethics-2020-106314] [Medline: 32366705]
25. Dehmel S, Kenning P, Wagner G, Liedtke C, Micklitz H, Specht-Riemenschneider L. Die Wirksamkeit Der Corona-Warn-App Wird Sich Nur Im Praxistest Zeigen: Der Datenschutz Ist Nur Eine Von Vielen Herausforderungen [The Effectiveness of the Corona-Warn-App Will Only Be Shown in Practical Tests: Data Protection Is Only One of Many Challenges]. Berlin: Veröffentlichungen des Sachverständigenrats für Verbraucherfragen; 2020.

26. Berger C, Calabrese R. Some explorations in initial interaction and beyond: toward a developmental theory of interpersonal communication. Hum Commun Res 1975 Jan;1(2):99-112. [doi: 10.1111/j.1468-2958.1975.tb00258.x]
27. Redmond MV. Uncertainty Reduction Theory. In: English Technical Reports and White Papers. Ames, IA: Iowa State University; 2015:3.
28. Venkatesh V, Thong JYL, Chan FKY, Hu PJH. Managing citizens' uncertainty in e-government services: the mediating and moderating roles of transparency and trust. Inf Syst Res 2016 Mar;27(1):87-111. [doi: 10.1287/isre.2015.0612]
29. Grace R, Tham JCK. Adapting Uncertainty Reduction Theory for crisis communication: guidelines for technical communicators. J Business Tech Commun 2020 Sep 21;35(1):110-117. [doi: 10.1177/1050651920959188]
30. Abeler J, Bäcker M, Buermeyer U, Zillessen H. COVID-19 contact tracing and data protection can go together. JMIR mHealth uHealth 2020 Apr 20;8(4):e19359 [FREE Full text] [doi: 10.2196/19359] [Medline: 32294052]
31. Cho H, Ippolito D, Yu YW. Contact tracing mobile apps for COVID-19: privacy considerations and related trade-offs. arXiv Preprint posted online March 30, 2020. [FREE Full text]
32. Mozur P, Zhong R, Krolik A. In Coronavirus fight, China gives citizens a color code, with red flags. The New York Times. 2020 Mar 01. URL: https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html [accessed 2020-08-04]
33. O'Neill PH. No, coronavirus apps don't need 60% adoption to be effective. MIT Technology Review. 2020 Jun 05. URL: https://www.technologyreview.com/2020/06/05/1002775/covid-apps-effective-at-less-than-60-percent-download/ [accessed 2020-07-31]
34. Berger CR. Uncertain outcome values in predicted relationships Uncertainty Reduction Theory then and now. Hum Commun Res 1986 Sep;13(1):34-38. [doi: 10.1111/j.1468-2958.1986.tb00093.x]
35. Walker HJ, Bauer TN, Cole MS, Bernerth JB, Feild HS, Short JC. Is This How I Will Be Treated? Reducing Uncertainty through Recruitment Interactions. Acad Mange J 2013 Oct;56(5):1325-1347. [doi: 10.5465/amj.2011.0196]
36. Tidwell LC, Walther JB. Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: getting to know one another a bit at a time. Hum Commun Res 2002 Jul;28(3):317-348. [doi: 10.1111/j.1468-2958.2002.tb00811.x]
37. Flanagin AJ. Commercial markets as communication markets: uncertainty reduction through mediated information exchange in online auctions. N Media Soc 2016 Jun 30;9(3):401-423. [doi: 10.1177/1461444807076966]
38. Lind E, van den Bos K. When fairness works: toward a general theory of uncertainty management. Res Organizational Behav 2002 Jan;24:181-223. [doi: 10.1016/s0191-3085(02)24006-x]
39. Shi M, Jiang R, Hu X, Shang J. A privacy protection method for health care big data management based on risk access control. Health Care Manag Sci 2020 Sep;23(3):427-442. [doi: 10.1007/s10729-019-09490-4] [Medline: 31338637]
40. Dar AB, Lone AH, Zahoor S, Khan AA, Naaz R. Applicability of mobile contact tracing in fighting pandemic (COVID-19): issues, challenges and solutions. Comput Sci Rev 2020 Nov;38:100307 [FREE Full text] [doi: 10.1016/j.cosrev.2020.100307] [Medline: 32989380]
41. Laato S, Islam AKMN, Islam MN, Whelan E. What drives unverified information sharing and cyberchondria during the COVID-19 pandemic? Eur J Inf Syst 2020 Jun 07;29(3):288-305. [doi: 10.1080/0960085x.2020.1770632]
42. Van Bavel JJ, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
43. Lee JJ, Kang K, Wang MP, Zhao SZ, Wong JYH, O'Connor S, et al. Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and preventive behaviors: cross-sectional online study. J Med Internet Res 2020 Nov 13;22(11):e22205 [FREE Full text] [doi: 10.2196/22205] [Medline: 33048825]
44. Rauschnabel PA, He J, Ro YK. Antecedents to the adoption of augmented reality smart glasses: a closer look at privacy risks. J Business Res 2018 Nov;92:374-384. [doi: 10.1016/j.jbusres.2018.08.008]
45. Conway L, Woodard SR, Zubrod A. Social psychological measurements of COVID-19: coronavirus perceived threat, government response, impacts, and experiences questionnaires. PsyArXiv Preprint posted online April 7, 2020. [FREE Full text] [doi: 10.31234/osf.io/z2x9a]
46. Antheunis ML, Valkenburg PM, Peter J. Getting acquainted through social network sites: testing a model of online uncertainty reduction and social attraction. Comput Hum Behav 2010 Jan;26(1):100-109. [doi: 10.1016/j.chb.2009.07.005]
47. Kim YH, Kim DJ. A study of online transaction self-efficacy, consumer trust, and uncertainty reduction in electronic commerce transaction. 2005 Presented at: 38th Annual Hawaii International Conference on System Sciences; January 2005; Big Island, HI. [doi: 10.1109/HICSS.2005.52]
48. Tenney A, Kværner J, Gjerstad KI. Uncertainty in environmental impact assessment predictions: the need for better communication and more transparency. Impact Assess Project Appraisal 2006 Mar;24(1):45-56. [doi: 10.3152/147154606781765345]
49. Schnackenberg AK, Tomlinson EC. Organizational transparency. J Manage 2016 Jul 09;42(7):1784-1810. [doi: 10.1177/0149206314525202]
50. Holland D, Krause A, Provencher J, Seltzer T. Transparency tested: the influence of message features on public perceptions of organizational transparency. Public Relations Rev 2018 Jun;44(2):256-264. [doi: 10.1016/j.pubrev.2017.12.002]

51. Rawlins BR. Measuring the relationship between organizational transparency and employee trust. Public Relations J 2008;2(2):1-21.
52. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
53. Bélanger F, Carter L. Trust and risk in e-government adoption. J Strategic Inf Syst 2008 Jun;17(2):165-176. [doi: 10.1016/j.jsis.2007.12.002]
54. Siegrist M. The influence of trust and perceptions of risks and benefits on the acceptance of gene technology. Risk Anal 2000 Apr;20(2):195-203. [doi: 10.1111/0272-4332.202020] [Medline: 10859780]
55. Rousseau DM, Sitkin SB, Burt RS, Camerer C. Not so different after all: a cross-discipline view of trust. Acad Manage Rev 1998 Jul 01;23(3):393-404. [doi: 10.5465/Amr.1998.926617]
56. Corbitt BJ, Thanasankit T, Yi H. Trust and e-commerce: a study of consumer perceptions. Electronic Commerce Res Applications 2003 Sep;2(3):203-215. [doi: 10.1016/s1567-4223(03)00024-3]
57. McKnight DH, Choudhury V, Kacmar C. Developing and validating trust measures for e-commerce: an integrative typology. Inf Syst Res 2002 Sep;13(3):334-359. [doi: 10.1287/isre.13.3.334.81]
58. Koufaris M, Hampton-Sosa W. The development of initial trust in an online company by new customers. Inf Manage 2004 Jan;41(3):377-397. [doi: 10.1016/j.im.2003.08.004]
59. Moon MJ. Fighting COVID-19 with agility, transparency, and participation: wicked policy problems and new governance challenges. Public Adm Rev 2020 May 20:651-656 [FREE Full text] [doi: 10.1111/puar.13214] [Medline: 32836434]
60. Chanley V, Rudolph T, Rahn W. The origins and consequences of public trust in government: a time series analysis. Public Opin Q 2000;64(3):239-256. [doi: 10.1086/317987] [Medline: 11114267]
61. Keele L. Social capital and the dynamics of trust in government. Am J Political Sci 2007 Apr;51(2):241-254. [doi: 10.1111/j.1540-5907.2007.00248.x]
62. Albu OB, Flyverbom M. Organizational transparency: conceptualizations, conditions, and consequences. Business Soc 2016 Jul 13;58(2):268-297. [doi: 10.1177/0007650316659851]
63. Schnackenberg AK, Tomlinson E, Coen C. The dimensional structure of transparency: a construct validation of transparency as disclosure, clarity, and accuracy in organizations. Hum Relations 2020 Jun 27:001872672093331-001872672093333. [doi: 10.1177/0018726720933317]
64. Jahansoozi J. Organization - stakeholder relationships: exploring trust and transparency. J Manage Dev 2006 Dec;25(10):942-955. [doi: 10.1108/02621710610708577]
65. Horvath R, Katuscakova D. Transparency and trust: the case of the European Central Bank. Appl Economics 2016 May 21;48(57):5625-5638. [doi: 10.1080/00036846.2016.1181833]
66. Stewart KJ. Trust transfer on the World Wide Web. Organ Sci 2003 Feb;14(1):5-17. [doi: 10.1287/orsc.14.1.5.12810]
67. Lu Y, Yang S, Chau PY, Cao Y. Dynamics between the trust transfer process and intention to use mobile payment services: a cross-environment perspective. Inf Manage 2011 Dec;48(8):393-403. [doi: 10.1016/j.im.2011.09.006]
68. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. JMIR mHealth uHealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
69. Li X, Hess TJ, Valacich JS. Why do we trust new technology? A study of initial trust formation with organizational information systems. J Strategic Inf Syst 2008 Mar;17(1):39-71. [doi: 10.1016/j.jsis.2008.01.001]
70. Aggelidis VP, Chatzoglou PD. Using a modified technology acceptance model in hospitals. Int J Med Inform 2009 Feb;78(2):115-126. [doi: 10.1016/j.ijmedinf.2008.06.006] [Medline: 18675583]
71. Bhattacherjee A, Sanford C. Influence processes for information technology acceptance: an elaboration likelihood model. MIS Q 2006;30(4):805-825. [doi: 10.2307/25148755]
72. Hsu C, Lin JC. Acceptance of blog usage: the roles of technology acceptance, social influence and knowledge sharing motivation. Inf Manage 2008 Jan;45(1):65-74. [doi: 10.1016/j.im.2007.11.001]
73. Bettiga D, Lamberti L, Lettieri E. Individuals' adoption of smart technologies for preventive health care: a structural equation modeling approach. Health Care Manag Sci 2020 Jun;23(2):203-214. [doi: 10.1007/s10729-019-09468-2] [Medline: 30684067]
74. van den Putte B, Yzer MC, Brunsting S. Social influences on smoking cessation: a comparison of the effect of six social influence variables. Prev Med 2005 Jul;41(1):186-193. [doi: 10.1016/j.ypmed.2004.09.040] [Medline: 15917010]
75. Gerrits JH, O'Hara RE, Piko BF, Gibbons FX, de Ridder DTD, Keresztes N, et al. Self-control, diet concerns and eater prototypes influence fatty foods consumption of adolescents in three countries. Health Educ Res 2010 Dec;25(6):1031-1041. [doi: 10.1093/her/cyq055] [Medline: 20861154]
76. Gorsuch RL, Butler MC. Initial drug abuse: a review of predisposing social psychological factors. Psychol Bull 1976 Jan;83(1):120-137. [Medline: 1019281]
77. Nicolaou AI, McKnight DH. Perceived information quality in data exchanges: effects on risk, trust, and intention to use. Inf Syst Res 2006 Dec;17(4):332-351. [doi: 10.1287/isre.1060.0103]
78. Ajzen I. The theory of planned behavior. Organizational Behav Hum Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]

79. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
80. King WR, He J. A meta-analysis of the technology acceptance model. Inf Manage 2006 Sep;43(6):740-755. [doi: 10.1016/j.im.2006.05.003]
81. Corona-Warn-App: Die wichtigsten Fragen und Antworten [Corona-Warn-App: The most important questions and answers]. Bundesregierung. 2020. URL: https://www.bundesregierung.de/breg-de/themen/corona-warn-app/corona-warn-app-faq-1758392 [accessed 2020-08-11]
82. Corona-Warn-App. GitHub. 2020. URL: https://github.com/corona-warn-app [accessed 2020-08-06]
83. Kennzahlen zur Corona-Warn-App [Key figures of the Corona-Warn-App]. Robert Koch-Institut. 2020. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Kennzahlen.pdf?__blob=publicationFile [accessed 2020-08-06]
84. Infektionsketten digital unterbrechen mit der Corona-Warn-App [Interrupt infection chains digitally with the Corona-Warn-App]. Robert Koch-Institut. 2020. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Warn_App.html [accessed 2020-06-24]
85. Open-Source-Projekt für Corona-Warn-App [Open source project for Corona-Warn-App]. 2020. URL: https://www.coronawarn.app/de/ [accessed 2020-06-24]
86. Leiner DJ. Too fast, too straight, too weird: post hoc identification of meaningless data in internet surveys. Survey Research Methods 2019;13(3):229-248. [doi: 10.18148/srm/2018.v13i3.7403]
87. Statistisches Jahrbuch 2019: Deutschland und Internationales [Statistical yearbook 2019: Germany and international affairs]. Statistisches Bundesamt. 2019. URL: https://www.destatis.de/DE/Themen/Querschnitt/Jahrbuch/statistisches-jahrbuch-2019-dl.pdf;jsessionid=BE46D99F4F1B92FD0BD4A58FCBA23DDC.internet8731?__blob=publicationFile [accessed 2020-08-06]
88. Ausstattung privater Haushalte mit Informations- und Kommunikationstechnik - Deutschland [Equipment of private households with information and communication technology - Germany]. Statistisches Bundesamt. 2020. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Ausstattung-Gebrauchsgueter/Tabellen/liste-infotechnik-d.html [accessed 2020-08-04]
89. Featherman MS, Pavlou PA. Predicting e-services adoption: a perceived risk facets perspective. Int J Hum-Comput Stud 2003 Oct;59(4):451-474. [doi: 10.1016/S1071-5819(03)00111-3]
90. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]
91. Hair JF, Hult GT, Ringle C, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) 2nd edition. Los Angeles: Sage Publications; Apr 05, 2016.
92. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Marketing Res 1981 Feb;18(1):39-50. [doi: 10.2307/3151312]
93. Venkatesh V, Thong J, Xu X. Consumer acceptance and use of information technology: extending the Unified Theory of Acceptance and use of technology. MIS Q 2012;36(1):157-178. [doi: 10.2307/41410412]
94. Goss-Sampson M. Statistical analysis in Jasp: a guide for students. static.jasp-stats.org. 2018. URL: https://static.jasp-stats.org/Statistical%20Analysis%20in%20JASP%20-%20A%20Students%20Guide%20v1.0.pdf [accessed 2020-07-15]
95. Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Soft 2012;48(2):1. [doi: 10.18637/jss.v048.i02]
96. Hooper D, Coughlan J, Mullen M. Structural equation modeling: guidelines for determining model fit. Electronic J Business Res Methods 2008;6(1):53-60.
97. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling Multidisciplinary J 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
98. Kline RB. Principles and Practice of Structural Equation Modeling. New York: The Guilford Press; 2005.
99. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
100. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [doi: 10.1037//0022-3514.51.6.1173] [Medline: 3806354]
101. Wu K, Zhao Y, Zhu Q, Tan X, Zheng H. A meta-analysis of the impact of trust on technology acceptance model: investigation of moderating influence of subject and context type. Int J Inf Manage 2011 Dec;31(6):572-581. [doi: 10.1016/j.ijinfomgt.2011.03.004]
102. Sheeran P, Webb TL. The intention-behavior gap. Soc Pers Psychol Compass 2016 Sep 05;10(9):503-518. [doi: 10.1111/spc3.12265]
103. Goldfarb A, Tucker C. Shifts in privacy concerns. Am Econ Rev 2012 May 01;102(3):349-353. [doi: 10.1257/aer.102.3.349]
104. Han PKJ, Klein WMP, Arora NK. Varieties of uncertainty in health care: a conceptual taxonomy. Med Decis Making 2011;31(6):828-838 [FREE Full text] [doi: 10.1177/0272989x11393976] [Medline: 22067431]

105. Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M. Features and functionalities of smartphone apps related to COVID-19: systematic search in app stores and content analysis. J Med Internet Res 2020 Aug 25;22(8):e20334 [FREE Full text] [doi: 10.2196/20334] [Medline: 32614777]
106. Morley J, Cowls J, Taddeo M, Floridi L. Ethical guidelines for COVID-19 tracing apps. Nature 2020 Jun;582(7810):29-31. [doi: 10.1038/d41586-020-01578-0] [Medline: 32467596]
107. Rivis A, Sheeran P. Descriptive norms as an additional predictor in the theory of planned behaviour: a meta-analysis. Curr Psychol 2003 Sep;22(3):218-233. [doi: 10.1007/s12144-003-1018-2]

Abbreviations

CB-SEM: covariance-based structural equation modeling
H: hypothesis
RQ: research question
SEM: structural equation modeling
TAM: technology acceptance model
URT: Uncertainty Reduction Theory
UTAUT: unified theory of acceptance and use of technology

Edited by G Eysenbach; submitted 19.11.20; peer-reviewed by P Banik, N Martin, C Jacob; comments to author 31.12.20; revised version received 06.01.21; accepted 16.01.21; published 08.02.21.

Please cite as:
Oldeweme A, Märtins J, Westmattelmann D, Schewe G
The Role of Transparency, Trust, and Social Influence on Uncertainty Reduction in Times of Pandemics: Empirical Study on the Adoption of COVID-19 Tracing Apps
J Med Internet Res 2021;23(2):e25893
URL: https://www.jmir.org/2021/2/e25893
doi:10.2196/25893
PMID:33465036

Original Paper

Suitability of Video Consultations During the COVID-19 Pandemic Lockdown: Cross-sectional Survey Among Norwegian General Practitioners

Tor Magne Johnsen[1,2*], MD; Børge Lønnebakke Norberg[1,2*], MD; Eli Kristiansen[1], MSc; Paolo Zanaboni[1,3], PhD; Bjarne Austad[2], MD, PhD; Frode Helgetun Krogh[1,2], MSc, MPBA; Linn Getz[2], MD, PhD

[1]Norwegian Centre for E-health Research, Tromsø, Norway

[2]Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway

[3]Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

[*]these authors contributed equally

Corresponding Author:
Eli Kristiansen, MSc
Norwegian Centre for E-health Research
Sykehusveien 23
Tromsø, 9019
Norway
Phone: 47 97162311
Email: eli.kristiansen@ehealthresearch.no

Abstract

Background: The COVID-19 pandemic imposed an acute, sharp rise in the use of video consultations (VCs) by general practitioners (GPs) in Norway.

Objective: This study aims to document GPs' experiences with the large-scale uptake of VCs in the natural experiment context of the pandemic.

Methods: A nationwide, cross-sectional online survey was conducted among Norwegian GPs during the pandemic lockdown (April 14-May 3, 2020). Each respondent was asked to evaluate up to 10 VCs. Basic demographic characteristics of the GPs and their practices were collected. The associations between GPs' perceived suitability of the VCs, the nature of the patients' main problems, prior knowledge of the patients (relational continuity), and follow-up of previously presented problems (episodic continuity) were explored using descriptive statistics, diagrams, and chi-square tests.

Results: In total, 1237 GPs (26% of the target group) responded to the survey. Among these, 1000 GPs offered VCs, and 855 GPs evaluated a total of 3484 VCs. Most GPs who offered VCs (1000/1237; 81%) had no experience with VCs before the pandemic. Overall, 51% (1766/3476) of the evaluated VCs were considered to have similar or even better suitability to assess the main reason for contact, compared to face-to-face consultations. In the presence of relational continuity, VCs were considered equal to or better than face-to-face consultations in 57% (1011/1785) of cases, as opposed to 32% (87/274) when the patient was unknown. The suitability rate for follow-up consultations (episodic continuity) was 61% (1165/1919), compared to 35% (544/1556) for new patient problems. Suitability varied considerably across clinical contact reasons. VCs were found most suitable for anxiety and life stress, depression, and administrative purposes, as well as for longstanding or complex problems that normally require multiple follow-up consultations. The GPs estimate that they will conduct about 20% of their consultations by video in a future, nonpandemic setting.

Conclusions: Our study of VCs performed in general practice during the pandemic lockdown indicates a clear future role for VCs in nonpandemic settings. The strong and consistent association between continuity of care and GPs' perceptions of the suitability of VCs is a new and important finding with considerable relevance for future primary health care planning.

(J Med Internet Res 2021;23(2):e26433) doi:10.2196/26433

KEYWORDS

video consultations; digital consultations; eHealth; general practice; primary health care; continuity of care; physician-patient relationship; patient safety; COVID-19; pandemic; telehealth; telemedicine; consultation; safety; cross-sectional; online survey

Introduction

Background

The digitalization of medical consultations in general practice (family medicine) has in many countries received increasing interest in recent years, with a particular focus on video consultations (VCs) [1-3]. However, the implementation of VCs by general practitioners (GPs) has been relatively slow and supported by limited and inconclusive evidence [4]. The emergence of the COVID-19 pandemic introduced an abrupt and strong stimulus for rapid adoption of VCs in many contexts, producing an effect that was arguably most striking in settings where a digitalization process had already started [5,6].

Prepandemic research on VCs in general practice has mostly been characterized by small studies on selected patient groups [7,8]. Results indicate that VCs might be useful for selected patients or health problems and have the potential for increased patient empowerment, practical convenience, and efficiency gains [9,10]. However, concerns have been raised regarding the clinical quality and suitability of VCs, and both patients and clinicians still consider face-to-face consultations the gold standard [7,8,11,12]. More knowledge is needed about the optimal use of VCs in general practice, from both an organizational and a clinical perspective [13].

In the context of face-to-face consultations, continuity of care [14] has been associated with positive health outcomes for patients, including increased life expectancy [15,16]. There is little knowledge regarding the impact of VCs on the doctor-patient relationship and to what extent a pre-existing doctor-patient relationship might impact the quality and outcome of VCs. The introduction of VCs in situations where GPs deliver continuity of care to their patients (relational continuity) and are familiar with their ongoing health problems (episodic continuity) may have a positive impact compared to situations where continuity of care is not established [17].

Use of VCs in Norway Before and During the COVID-19 Pandemic

Before the COVID-19 pandemic, around 3% of all GP consultations in Norway were performed digitally [18]. The societal lockdown in Europe in spring 2020 had a strong impact on general practice in Norway. On March 16, 2020, the Norwegian Ministry of Health and Care Services encouraged all GPs to adopt a solution for VCs [19]. The reimbursement system for GPs was also temporarily modified to strengthen the use of medical consultations via text, video, or telephone (Textbox 1). Data from the main providers of VC solutions showed that, while almost all 4822 GPs in Norway installed a solution for VCs, less than 2000 GPs used VCs during the period April 15-May 3, 2020.

During the first phase of the lockdown in mid-March 2020, many GP offices restricted physical access and triaged all contacts via telephone or online communication. Almost 60% (171,169/299,148) of all GP consultations in Norway were performed digitally from March 16 to March 22 (Multimedia Appendix 1). In the last week of March, 25% (34,814/141,501) of all digital consultations were VCs (Multimedia Appendix 2). From May 11 to May 17, the proportion of digital consultations decreased to 30% (85,026/286,419), of which 19% (16,278/85,026) were VCs.

Most GP offices still offered face-to-face consultations for urgent issues. However, digital consultations were conducted even in situations where a physical examination would have been deemed necessary before the COVID-19 lockdown, such as acute abdominal pain or chest pain. At the time of data collection, the main technical solution for VCs in Norway was an external video application not integrated with the GP's electronic patient record systems. However, the patients' medical record was available to the GPs during the VCs.

Textbox 1. The Norwegian General Practitioner Scheme.

The Norwegian health care system is based on the principles of universal access, decentralization, and continuity of care [20]. Since 2001, all Norwegian citizens may sign up with (and change, if desired) a GP, and 99% have chosen to do so. The system is financed by taxation, together with income-related employee and employer contributions and out-of-pocket payments (copayments). Private medical insurance is limited. Although national health care policy is controlled centrally, responsibility for the provision of primary health care is decentralized. GPs act as coordinators of municipal services and gatekeepers to specialized care. On average, a GP has about 1100 patients and often provides other medical services in the municipality one day per week.

Evaluation of VCs During the COVID-19 Pandemic

In association with the COVID-19 pandemic, recommendations regarding the use of VCs by GPs and their patients have been issued based on clinical expertise and relevant evidence [13,21]. The lockdown led to a rapid uptake of VCs during a very short time period, creating a natural experiment where the effectiveness and suitability of VCs could be explored across a wide range of health problems [5].

Ideally, the large-scale implementation of VCs should have been rigorously monitored by detailed research on both GPs and patients. Due to the pressing circumstances of the lockdown, such systematic evaluation was not deemed feasible. However, conducting a large-scale survey of GPs' experiences with the use of VCs during the lockdown was achievable and can contribute to filling important knowledge gaps.

The overall aim of this survey was to explore how GPs in Norway perceived the suitability of VCs compared to ordinary face-to-face consultations during the COVID-19 lockdown. In addition to addressing the suitability of VCs across a wide range of health problems (reasons for contact), we were also interested in knowing whether continuity of care (ie, prior knowledge of the patient/problem) had an impact on GPs' perceptions of VC suitability.

Methods

Study Design and Setting

A prospective nationwide online cross-sectional survey was conducted among GPs in Norway during the COVID-19 pandemic lockdown (April 14-May 3, 2020). The survey was addressed to all GPs registered in Norway. Basic demographic characteristics of the GPs and their practices were collected. Each GP also indicated the number of consultations and other activities conducted during the day when the survey was taken.

A central part of the survey addressed GPs' experiences with VCs before and during the pandemic. Each GP was asked to evaluate up to 10 consecutive (or otherwise unselected) VCs during the COVID-19 lockdown, preferably conducted during the same day. The evaluation of each VC included 13 questions that covered the GPs' prior knowledge of the patient (relational continuity), whether the reason for contact was a new problem or a follow-up (episodic continuity), the total number of presented problems, the nature of the main problem (as perceived by the GP), the perceived suitability of VC compared to an envisaged face-to-face consultation for the main problem, and actions (one or more) taken by the GP during/after the VC. The GP's perception of the patient's satisfaction with the VC, the technical quality of each VC, and their willingness to use VC in a similar situation after the COVID-19 pandemic were also recorded. Finally, the GPs were asked to estimate the overall proportion of VCs they personally envisaged in their practice in a "normalized" future, in light of their accumulated experience with the medium.

Questions were multiple choice with 2-11 alternative answers, depending on the topic. Questions concerning users' experiences were scored on a 3-point or 4-point Likert scale. Regarding patients' reasons for contact, a total of 78 alternatives were offered. This list was informed by the International Classification of Primary Care, second edition (ICPC-2), but was less detailed. Regarding actions taken during or after each VC, 16 alternatives were offered. We consulted the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) to develop the survey and report its results [22]. The survey was pilot tested by a panel of experienced GPs. The survey was conducted in Norwegian through the Netigate application. The results have been translated into English for the purpose of publication.

Data Collection

To obtain access to GPs' updated contact emails on very short notice, the research team collaborated with Norwegian Health Informatics, a web-based portal that hosts an online clinical decision support product (NEL), to which approximately 98% of all Norwegian GPs subscribe [23]. An invitation was sent to all subscribers by a unique link that ensured both the authenticity and anonymity of the respondents. GPs who did not receive a personal invitation by email or were nonsubscribers of NEL were encouraged by a well-established social media group for GPs to register their email addresses on the Norwegian Health Informatics website. It took approximately 30-60 minutes to complete the survey. Data collection was undertaken in the period April 14-May 3, 2020. Several reminders were sent by email and social media.

Data Analysis

Results were summarized by descriptive statistics, diagrams, and chi-square tests with 95% CIs. Background data from this survey were compared to available information on all GPs in Norway. The associations between suitability of VCs and relational continuity, episodic continuity, and the nature of the patients' main problems were explored by diagrams, tables, and chi-square tests. The defined significance level was .05.

Before further analysis was performed, the 78 alternative reasons for contact were merged into 27 more overarching categories (eg, knee, hip, shoulder, and back problems were merged into "musculoskeletal issues"). We also merged "better" and "same" into one category regarding suitability. Data were analyzed in IBM Statistical Package for the Social Sciences (SPSS; Version 26.0, IBM Corp).

Ethics

Ethical considerations were included in all phases of the survey. Participating GPs were informed that participation was voluntary and anonymous. We did not elicit sensitive information or demographic characteristics that could reveal the identity of the GPs. For the evaluated VCs, we did not elicit patients' age, sex, specific diagnoses, or other sensitive or person-related information. Distribution of the survey to GPs' email addresses was handled by an independent party (Norwegian Health Informatics). No linkage key was established, and participants' IP numbers were not accessible to any party. Further approvals were thereby not required, according to Norwegian health research legislation, verified by the Norwegian Centre for Research Data (NSD).

Results

Characteristics of the GPs

A total of 1237 GPs participated in the survey, representing 26% (1237/4822) of the total GP population in Norway [24]. Of these, 1000 (81%) answered that they were equipped to offer VCs at the time of the survey, and 855 contributed evaluations of at least one VC (Table 1).

On average, each GP conducted 20 consultations during the surveyed working day. Of these, 6.9 (34.5%) were face-to-face consultations, 5.3 (26.5%) were VCs, 3.3 (16.5%) were text-based e-consultations, and 4.5 (22.5%) were telephone consultations. Of the 855 participants, 74 (9%) had no face-to-face consultations on the study day. Most of the respondents (80%; 687/855) did not have any experience with the use of VCs before the COVID-19 pandemic.

Table 1. Characteristics of the 855 general practitioners who evaluated one or more video consultations.

Characteristics of the general practitioners (N=855)	Participants, n (%)
Gender	
Female	480 (56.2)
Male	368 (43.0)
No answer	7 (0.8)
Experience in years as a general practitioner	
0-5	174 (20.3)
6-10	189 (22.1)
11-20	287 (33.6)
>20	205 (24.0)
Inhabitants of the municipality of practice	
<10,000	130 (15.2)
10,000-50,000	295 (34.5)
50,000-100,000	143 (16.7)
100,000-500,000	217 (25.4)
>500,000	70 (8.2)
Experience with video consultations before the COVID-19 pandemic	
None	687 (80.3)
Limited (1-50 video consultations)	123 (14.4)
Relatively good (>50 video consultations)	45 (5.3)

Description of the VCs

On average, each GP provided an evaluation of 3.8 VCs. The final data set included 3484 unique VCs between GPs and patients (Table 2).

In most VCs (79%; 2760/3484), the GP knew the patient well beforehand, while the GPs described only 8% (276/3484) of the patients as previously unknown. More than half of the consultations (55%; 1921/3481) were a follow-up of a previous problem. On average, 1.9 (median 2.0) problems/issues were discussed during each consultation. Half of the VCs (51%; 1766/3476) were considered to have similar or even better suitability to assess the main reason for contact compared to face-to-face consultations. For 15% (514/3476) of the VCs, the GPs expressed concern that they might not have detected potential signs of serious illness. The lack of opportunity to physically examine the patient was reported as a “major loss” or “some loss” in 25% (884/3475) and 36% (1232/3475) of the VCs, respectively. In 85% (2967/3475) of cases, the GPs perceived that the patient was satisfied with the VC. The technical quality was considered good in 90% (3118/3475) of the VCs. Half of the GPs (1704/3475) considered it realistic to handle a similar issue by VC after the COVID-19 pandemic.

Table 2. Characteristics of the 3484 recorded video consultations. Missing answers (ranging from 0-9 general practitioners per question) are not displayed.

Characteristics of the video consultations (N=3484)	Values, n (%)
General practitioner's pre-existing knowledge of the patient	
Very good	1788 (51.3)
Good	972 (27.9)
Some	448 (12.9)
None	276 (7.9)
Main reason for contact	
New problem	1560 (44.8)
Follow-up	1921 (55.1)
Total number of contact reasons discussed	
1	1471 (42.2)
2-4	1930 (55.4)
>4	75 (2.2)
Suitability of video consultation compared to a face-to-face consultation for the same reason	
Better or same	1766 (50.7)
Worse	1709 (49.1)
Suitability of video consultation to assess the severity of the main reason for contact compared to a face-to-face consultation	
Better or same	1767 (50.7)
Worse	1709 (49.1)
Loss from not being able to examine the patient physically	
No loss	1359 (39.0)
Some loss	1232 (35.4)
Major loss	884 (25.4)
Concern about not picking up signs of serious illness	
Not worried	2009 (57.7)
Neutral	953 (27.3)
Worried	514 (15.0)
General practitioner's perception of patient satisfaction with video consultation	
Very satisfactory	988 (28.4)
Satisfactory	1979 (56.8)
Unsatisfactory	368 (10.6)
Do not know	140 (4.0)
General practitioner's satisfaction with technology (connection, sound, image)	
Very satisfactory	1433 (41.1)
Satisfactory	1685 (48.4)
Unsatisfactory	273 (7.8)
Video consultation terminated due to technical problems	84 (2.4)
Motivation to conduct a video consultation for a similar health problem (reason for contact) in a nonpandemic future	
Yes	1704 (48.9)
Do not know	768 (22.0)
No	1003 (28.8)

Suitability of VCs and Continuity of Care

The association between GPs' perceptions of the suitability of VCs and their previous knowledge of their patients (relational continuity) is presented in Figure 1. When the GPs knew their patients "very well" beforehand, VCs were considered as having "better" or "same" suitability, compared to face-to-face consultations in 57% (1011/1785) of cases. When the patient was "unknown," the corresponding suitability rate dropped to 32% (87/274). The difference in proportions of better/same and worse suitability of VCs between the groups of relational continuity was statistically significant (χ^2_3=105.3, *P*<.001).

In Figure 2, we present the association between the GPs' perceptions of suitability of VCs and their previous knowledge of a given patient's presented problem (episodic continuity). VCs were considered better/same compared to envisaged face-to-face consultations in 61% (1165/1919) of follow-up consultations, as opposed to 35% (544/1556) when the patient presented a new problem. The difference in proportions of better/same and worse suitability of VCs between the groups of episodic continuity was statistically significant (χ^2_1=227.9, *P*<.001).

In Table 3**,** we have combined relational and episodic continuity of care. We present the proportion of VCs where the suitability is considered better or the same, compared to envisaged face-to-face consultations for the same contact reason. When the GP's prior knowledge of a patient was "very good" or "good" and the problem was a "follow-up," 62% (1070/1719) of the VCs were considered equally or better suited than envisaged face-to-face consultations. The corresponding proportion drops to 30% (155/521) when the GP's prior knowledge of the patient was "some" or "none" and the problem was "new." The differences in proportion of the suitability of VCs in Table 3 are statistically significant (*P*<.001).

Figure 1. Association between relational continuity (previous knowledge of the patient) and general practitioners' perceived suitability of video consultations compared to an envisaged face-to-face consultation for the same issue (95% CIs are illustrated by lines). GP: general practitioner; VC: video consultation.

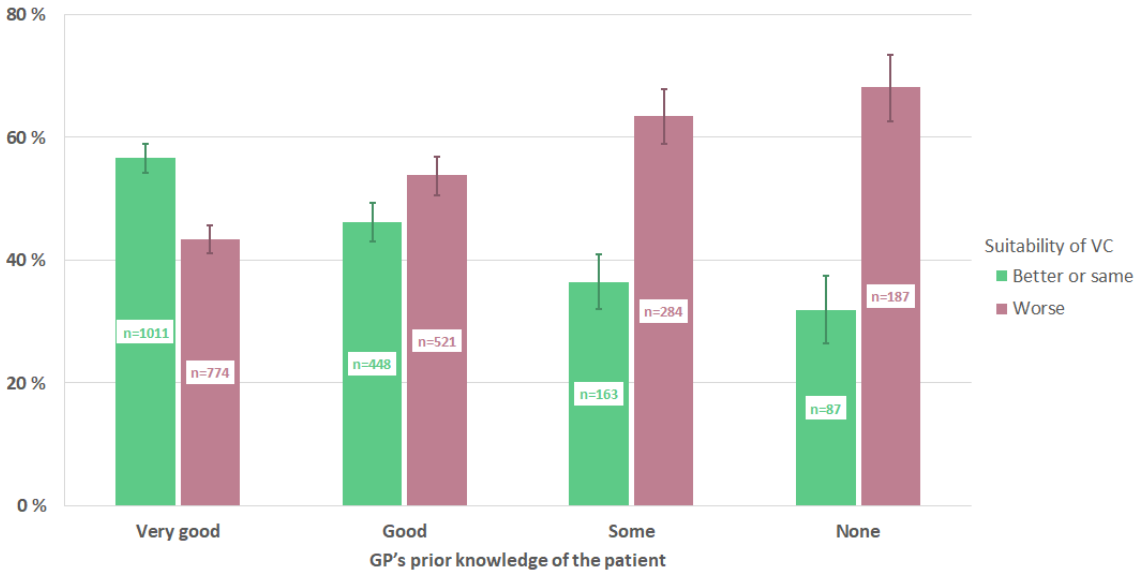


Figure 2. Association between episodic continuity (new problem or follow-up of previously defined problem) and general practitioners' perceived suitability of the video consultations, compared to envisaged face-to-face consultations for the same problem (95% CIs are illustrated by lines). VC: video consultation.

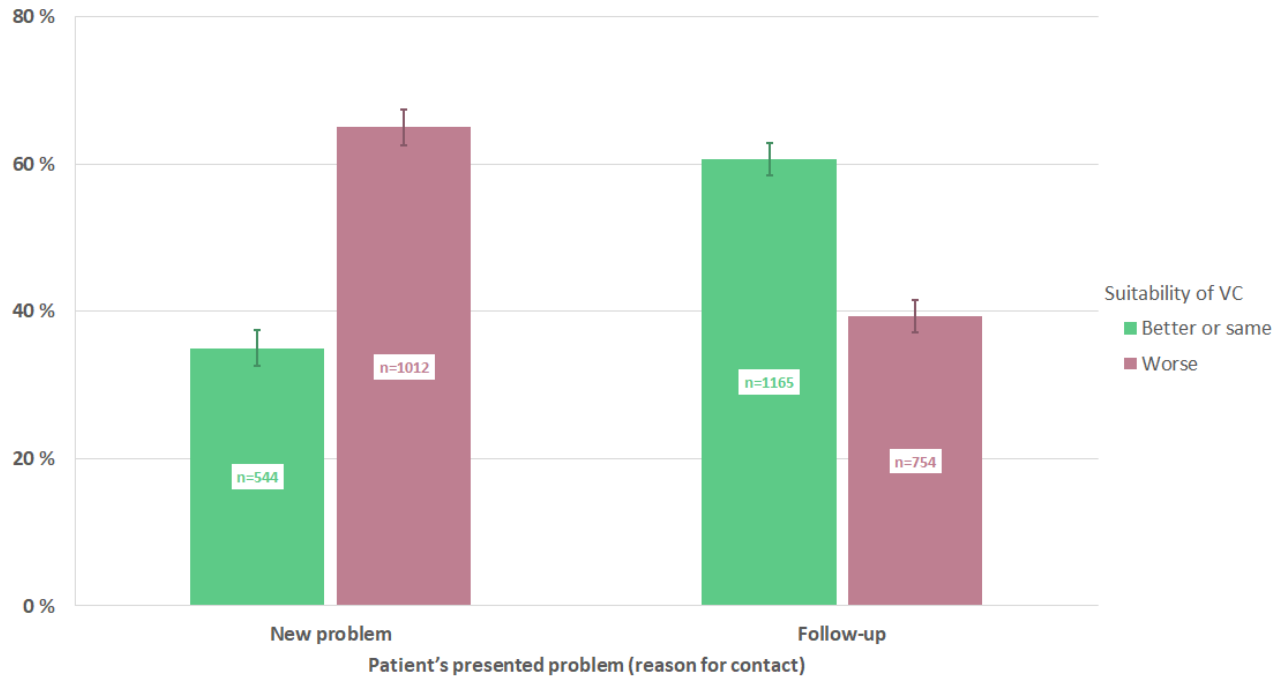


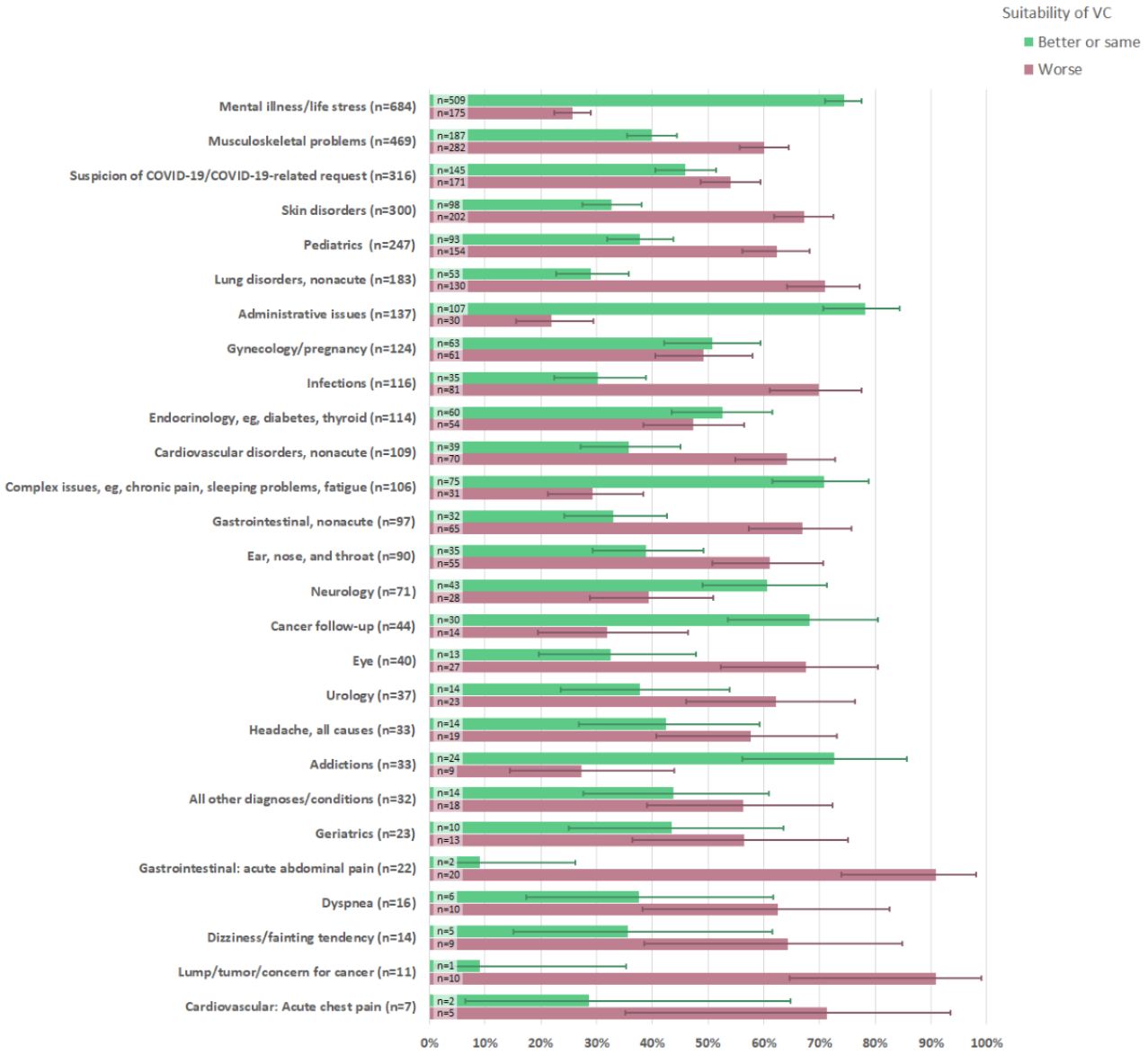
Table 3. Video consultations perceived by the GPs as equally or better suited compared to an envisaged face-to-face consultation, shown as combinations of pre-existing knowledge of the patient (relational continuity) and the main reason for contact (episodic continuity).

Main reason for contact	Pre-existing knowledge of patient			
	Very good/good (n=2754)		Some/none (n=721)	
	Number of video consultations, n/N	Proportion, % (95% CI)	Number of video consultations, n/N	Proportion, % (95% CI)
New problem (n=1556)	389/1035	38 (35-41)	155/521	30 (26-34)
Follow-up (n=1919)	1070/1719	62 (60-65)	95/200	48 (41-54)

Suitability of VCs and Reasons for Contact

GPs considered VCs to be equally or better suited for several reasons for contact, compared to envisaged face-to-face consultations (Figure 3); examples include mental illness/life stress (509/684; 74%, 95% CI 71%-78%) and various administrative purposes (107/137; 78%, 95% CI 70%-84%). On the other hand, other issues, including musculoskeletal problems (187/469; 40%, 95% CI 36%-44%) and skin disorders (98/300; 33%, 95% CI 28%-38%), were regarded as less suitable for VCs. VCs were also considered less suitable in situations involving acute chest pain, stomach pain, and fear/investigation of a potential new cancer.

Figure 3. General practitioners' perceived suitability of video consultations compared to an envisaged face-to-face consultation for the same issue, according to the nature of the main problem/reason for contact. The contact reasons are presented in decreasing frequency (n) from top to bottom, and 95% CIs are illustrated by lines. VC: video consultation.



Furthermore, for each main reason for contact, the perceived suitability of VCs differed according to episodic continuity of care (Table 4). For instance, the suitability of VCs for skin disorders was 30% (69/234) for a new problem and 44% (29/66) for a previously discussed problem. Corresponding numbers were 18% (33/183) and 54% (154/286) for musculoskeletal problems, and 20% (4/20) and 77% (39/51) for neurology disorders. The impact of relational continuity of care on the suitability of VCs for individual health problems is not reported due to low statistical power (few unknown patients).

Table 4. Suitability of video consultations compared to face-to-face consultations, association with main reason for contact (new problem/follow up), and the 20 most common issues/presented problems, N=3350.

Issue/presented problem	Main reason for contact							
	New problem (n=1484)				Follow-up (n=1866)			
	Better or same (n=529)		Worse (n=955)		Better or same (n=1140)		Worse (n=726)	
	Contacts, n/N	Proportion, % (95% CI)	Contacts, n/N	Proportion, % (95% CI)	Contacts, n/N	Proportion, % (95% CI)	Contacts, n/N	Proportion, % (95% CI)
Mental illness/life stress	67/105	64 (54-73)[a]	38/105	36 (28-46)[a]	442/579	76 (73-80)[a]	137/579	24 (20-27)[a]
Musculoskeletal problems	33/183	18 (13-24)[a]	150/183	82 (76-87)[a]	154/286	54 (48-60)	132/286	46 (40-52)
Suspicion of COVID-19 or COVID-19–related	99/222	45 (38-51)	123/222	55 (49-62)	46/94	49 (39-59)	48/94	51 (41-61)
Skin disorders	69/234	30 (24-36)[a]	165/234	71 (64-76)[a]	29/66	44 (32-56)	37/66	56 (44-68)
Children	52/173	30 (24-37)[a]	121/173	70 (63-76)[a]	41/74	55 (44-66)	33/74	45 (34-56)
Lung	27/96	28 (20-38)[a]	69/96	72 (62-80)[a]	26/87	30 (21-40)[a]	61/87	70 (60-79)[a]
Administrative issues	45/57	79 (67-88)[a]	12/57	21 (12-33)[a]	62/80	78 (68-86)[a]	18/80	23 (14-33)[a]
Gynecology and pregnancy	24/50	48 (35-62)	26/50	52 (38-65)	39/74	53 (41-64)	35/74	47 (36-59)
Infection	25/84	30 (21-40)[a]	59/84	70 (60-79)[a]	10/32	31 (17-48)	22/32	69 (52-83)
Endocrinology	6/15	40 (19-65)	9/15	60 (35-81)	54/99	55 (45-64)	45/99	46 (36-55)
Cardiovascular	2/37	19 (9-34)[a]	30/37	81 (66-91)[a]	32/72	44 (33-56)	40/72	56 (44-67)
Complex issues and disorders	10/20	50 (29-71)	10/20	50 (29-71)	65/86	76 (66-84)[a]	21/86	24 (16-34)[a]
Gastrointestinal	12/47	26 (15-39)[a]	35/47	75 (61-85)[a]	20/50	40 (27-54)	30/50	60 (46-73)
ENT (ear, nose, and throat)	23/68	34 (23-46)[a]	45/68	66 (54-77)[a]	12/22	55 (34-74)	10/22	46 (26-66)
Neurology	4/20	20 (7-41)[a]	16/20	80 (59-93)[a]	39/51	77 (64-86)[a]	12/51	24 (14-36)[a]
Cancer follow-up	2/3	67 (18-96)	1/3	33 (4-82)	28/41	68 (53-81)[a]	13/41	32 (19-47)[a]
Eye	12/36	33 (20-50)[a]	24/36	67 (51-80)[a]	1/4	25 (3-72)	3/4	75 (28-97)
Urology	9/22	41 (23-62)	13/22	59 (39-78)	5/15	33 (14-58)	10/15	67 (42-86)
Headache	1/8	13 (1-45)[a]	7/8	88 (55-99)[a]	13/25	52 (33-71)	12/25	48 (30-67)
Intoxication and addiction	2/4	50 (12-88)[a]	2/4	50 (12-88)[a]	22/29	76 (58-89)[a]	7/29	24 (12-42)[a]

[a]Nonoverlapping CIs.

Suitability of VCs and Actions Taken in the VCs

A total of 7647 actions were registered by the GP during/after the VCs (Table 5). The most common action (47%; 1646/3476) was "comprehensive advice and guidance," followed by planning a new VC or text-based e-consultation (34%; 1197/3476).

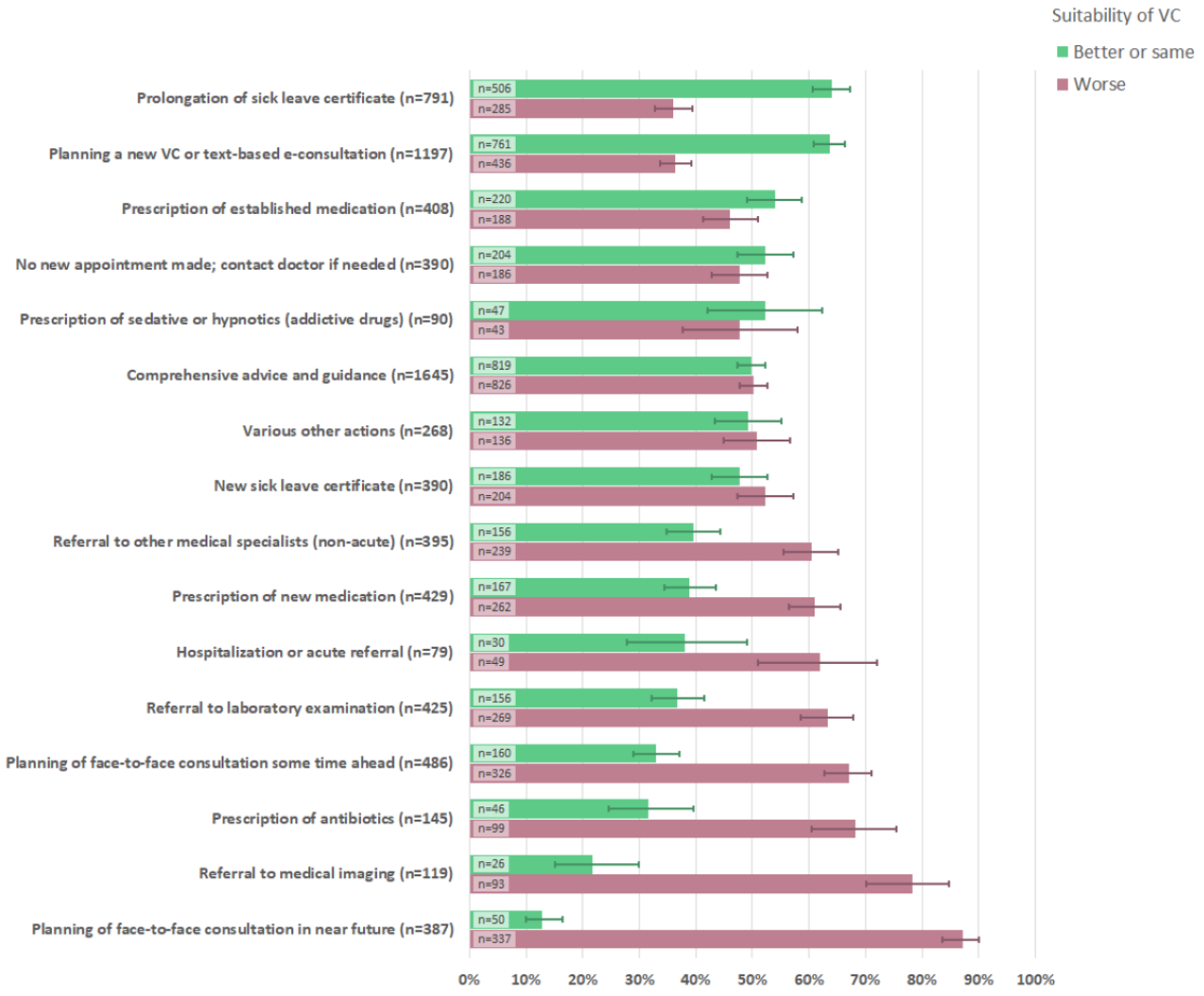
VCs were deemed to be less suitable in situations where a physical follow-up was subsequently planned (ie, a face-to-face consultation in the near future, referrals to medical imaging, laboratory examinations, and hospitalizations; Figure 4). Prescription of antibiotics or new medications were also situations where VCs were considered less suitable. Conversely, VCs were perceived as suitable for prolonging sick leaves.

Table 5. Action taken during/after the 3484 evaluated video consultations. Several actions could be registered for each video consultation.[a]

Type of actions taken during/after the video consultation (N=7647)	Count, n	%
Comprehensive advice and guidance	1646	47
Planning a new video consultation or text-based e-consultation	1197	34
Prolongation of sick leave certificate	791	23
Planning of face-to-face consultation some time ahead	486	14
Prescription of new medication	429	12
Referral to laboratory testing	425	12
Renewal of established medication	408	12
Referral to other medical specialists (nonacute)	395	11
New sick leave certificate	391	11
No new appointment made; contact doctor if needed	390	11
Planning of face-to-face consultation in near future	388	11
Various other actions	268	8
Prescription of antibiotics	145	4
Referral to medical imaging	119	3
Prescription of sedatives or hypnotics (addictive drugs)	90	3
Hospitalization or acute referral	79	2

[a]In total, 8 evaluated video consultations did not have a registered action.

Figure 4. Associations between general practitioners' perceived suitability of video consultations and actions taken during/after the consultation. The actions taken are sorted by general practitioners' perceived suitability (better or same) of the video consultation in question, displayed in decreasing order. 95% CIs are illustrated by lines. VC: video consultation.



Envisaging the Place for VCs in a Normalized Future

Based on their accumulated experience with VCs, participating GPs estimated that on average, 19% of all their consultations (median 20%) could be conducted via video in a normalized, nonpandemic future. Further analyses showed that this estimate was not associated with the GP's geographic context, previous experience with VCs, age, or gender.

Discussion

Principal Findings

This online, cross-sectional survey conducted among GPs in Norway during the COVID-19 pandemic lockdown provides new knowledge about GPs' experiences with large-scale uptake of VCs within the context of an established national primary health care system. Data collection took place in a phase of the lockdown when infection rates had peaked and most GPs had adapted their working practices, including implementation of VCs, to the pandemic circumstances.

Overall, VCs were perceived as equally or even more suitable than face-to-face consultations in about half of the 3484 evaluated cases. VCs appeared significantly more appropriate in the context of an established doctor-patient relationship and in relation to a previously defined reason for contact. GPs expressed concern about immediate patient safety (ie, the risk of missing signs of serious disease) in 15% (514/3476) of the evaluated cases.

Validating the Concept of Suitability

Most of the survey was developed around the perceived suitability of VCs in general practice, compared to ordinary face-to-face consultations. It was therefore essential to ensure that the participating GPs had a shared understanding of the concept of suitability. Face validity of the term was assessed when the survey was pilot tested and was found to be satisfactory. Moreover, GPs were instructed to evaluate the suitability of VCs without emphasis on the evident gain of reducing the risk of viral contamination. Furthermore, the analysis of the association between perceived suitability and actions taken in the VCs contributed to the validation of the

suitability concept. Consultations categorized with low suitability for VCs often had an "unfinished" character, thus requiring further clinical investigations and/or referrals that could have been more effectively managed in an ordinary face-to-face consultation (Figure 4).

Impact of Continuity of Care on the Perceived Suitability of VCs

A central finding in this survey was a strong and consistent association between both relational and episodic continuity of care and GPs' perceptions of the suitability of VCs (Figures 1 and 2, Tables 3 and 4). When GPs knew the patients very well beforehand, VCs were considered equally or better suited than face-to-face consultations in 57% (1011/1785) of cases, as opposed to 32% (87/274) when the patient was previously unknown. Moreover, the suitability rate for follow-up consultations was 61% (1165/1919), while for new problems it was 35% (544/1556). It is important to note that these findings arose within the Norwegian GP list system and that the patient was totally new to the GP in only 8% (276/3484) of the evaluated VCs. To our knowledge, our study presents the first large-scale analysis of the impact of continuity of care on GPs' perceptions regarding the clinical suitability of VCs. It remains to be seen whether the quantifiable, beneficial outcomes of continuity of care found in physical consultation settings can be replicated in relation to VCs. Beyond this question, our results are in good accordance with previous findings, deliberations, and recommendations regarding quality potentials and pitfalls related to VCs [10,21,25,26].

Suitability of VCs for Specific Contact Reasons

The predefined list of medical issues used to categorize the main contact reason (presented problem) for each evaluated VC was developed for the purpose of this survey to provide a clinically relevant overview while being general enough to safeguard patient anonymity. We acknowledge that the opportunity to categorize only one (perceived by the GP as the main) reason for contact underrates the clinical complexity of the evaluated VCs, as many of them dealt with two or more health problems (Table 2). In our VC material, mental problems and life stress, musculoskeletal disorders, and COVID-19–related contacts occurred most frequently. When most countries in Europe imposed a lockdown in March 2020, VCs were predicted to be potentially useful for consulting about COVID-19 for people with heightened anxiety, mild symptoms suggestive of COVID-19, or more severe symptoms [13]. The results from our survey confirm these predictions.

Moreover, under the given circumstances, we found that VCs were suitable for a variety of mental health problems, along with other chronic or complex issues such as chronic pain, tiredness, sleeping problems, follow-up of established cancer treatment, and administrative purposes (Figure 3). These results are also in line with previous expectations that VCs could replace in-person visits for contacts related to chronic disease reviews, counselling, or other talking therapies, or administrative appointments such as sick leave certificates [13,27]. Chronic and complex problems are prevalent in general practice and typically involve several consultations with the GP over time [28]. In the presence of continuity of care, such problems can typically be handled in a collaborative dialogical process between doctor and patient. Further research might elucidate to what extent effective clinical relationships can be established and maintained through VCs.

Contacts related to skin problems, pediatric issues, and acute and potentially severe health problems (such as abdominal pain, chest pain, and respiratory difficulties) were found to be the least suitable for VCs. Real-time video might be a feasible alternative to face-to-face and store-and-forward (asynchronous) consultations for selected skin issues [29,30]. However, the evidence regarding the diagnostic accuracy of VCs is still weak [30]. Particular caution must be shown in situations where patient safety might be at risk, including the evaluation of potential malignancy [29-31]. Figure 3 provides an informative overall overview of GPs' perceived suitability of VCs for various contact reasons, but there is a clear need for more research to refine the knowledge about suitability for VCs in a normalized setting.

Our material included only 23 cases where the main problem was categorized as "geriatric," and the majority of these cases were classified with low suitability for VC. This aligns well with previous concerns that VCs may not be suitable for use by frail, older patients [8]. Overall, we believe that our results regarding clinical suitability are in good accordance with previous research and recommendations for VCs [13,26], as well as with a recent paper on the interpretative and contextualized nature of diagnostic "knowing" in general practice [32].

Characteristics of the GPs

Our study participants constitute approximately 26% (1237/4822) of all registered GPs in Norway. Their demographic characteristics are well representative of the GP population in Norway, with the exception of a slightly higher representation of younger doctors [24]. Since the survey was distributed through a unique link to each respondent, multiple responses from the same source can be ruled out.

Before deciding to participate in the survey, invited GPs were informed that VCs would be a central topic, but not a prerequisite for participation. This may nonetheless have attracted GPs with a positive attitude toward digital solutions. As mentioned in the introduction, many GPs who acquired VC equipment in relation to the pandemic lockdown hardly used it in practice. As previously argued, the implementation of VCs is not mainly a simple question of equipment installation, but a complex process of integrating a new consultation modality into established working routines [33]. As explained in the introduction, our estimates suggest that almost 50% of Norwegian GPs who used VCs in April 2020 participated in this survey, which is a clear strength.

Selection Mechanisms Behind the Evaluated VCs

On average, each participating GP performed five VCs on the day of the survey, while typically evaluating three of these. The survey instructed the GPs not to purposefully select particular types of VCs for evaluation. Although we cannot rule out possible selection bias, it is unlikely that GPs would deliberately

restrict their contribution to specific cases. The risk of substantial confirmation bias is therefore considered low.

However, other important selection mechanisms may have influenced our VC material. The study was undertaken in a lockdown period when patients were typically recommended to consult their GPs digitally or by phone before a physical consultation could be considered. In some GP offices, patients could book a VC directly through an online booking system. Beyond this, there were no formal guidelines on how to select patients for VCs. Irrespective of variations in booking systems, many GP offices would arrange physical consultations for selected urgent health problems. Nevertheless, our material includes numerous cases that would normally be considered poorly suited for VC (eg, acute abdomen and chest pain). This aligns well with the fact that participating GPs relatively often reported that they missed the opportunity to perform a physical examination (61% of cases) or expressed concern that they might have misjudged the severity of the presented health problem (15% of cases). Our results cannot be generalized to the use of VCs under nonpandemic circumstances, but we believe they have substantial external validity.

Current Status and Future Perspectives

The COVID-19 pandemic has led to a crisis for many people and organizations worldwide. At the same time, it has offered opportunities to rethink what is important in general practice and really put new consultation modalities to the test [34]. Overall, our study indicates that VCs have a definite role in future general practice. A crucial question is how to apply VCs to appropriate purposes in an organizationally sustainable manner. On the positive side, VCs can facilitate access and provide rapid solutions in well-selected situations. They can also enhance access to care for disadvantaged or vulnerable patients who are digitally literate but reluctant to visit the GP office. On the flip side, too liberal a use of digital consultations may contribute to a more transactional and less relational way of dealing with deeply human issues, potentially undermining the health-promoting potential of person-centered, longitudinal care [17,25]. Furthermore, it is important to monitor the impact of the implementation of digital consultation modalities on the total clinical workload [35,36]. Patients' thresholds for contacting the GP for self-limiting conditions may decrease when digital options are easily accessible. Ineffective selection of cases for VCs might trigger physical follow-up consultations for the same problem, leading to increased workloads without associated clinical benefit.

In our study, GPs experienced technical difficulties in 10% of the consultations, somewhat less than a comparable study performed in New York City in March 2020, where technical problems were reported in 13% of consultations [37]. This indicates that more seamless integration of VCs with GPs' electronic record systems still remains a priority. Despite some technical challenges, the GPs reported that 85% of the patients seemed satisfied with the VCs, in accordance with existing knowledge on patient experience with VCs [9,37,38]. In about half the consultations, the GPs in our survey also deemed it realistic to handle a similar issue by VC in a future, normalized situation. At first glance, this might reflect a strong belief in digitalized health care. However, our findings emerged in an extraordinary setting, where many patients and GPs reportedly felt gratitude for being able to "see" each other at all, despite the ominous threat of contagion and dramatic lockdown measures.

Comparison of clinical practice during lockdown with practice-as-usual reveals numerous discrepancies, both obvious and subtle [39]. Experience-based insight into these differences and their impact on the effectiveness and quality of care is likely to explain why the typical GP in our survey envisaged conducting only about 20% of consultations by video in a normalized future. This prediction is substantially lower than the level of enthusiasm otherwise reflected in our material but would nevertheless represent an important organizational leap for Norwegian general practice as a whole.

Conclusion

Our study of VCs performed in Norwegian general practice during the pandemic lockdown indicates a future role for VCs in future, nonpandemic settings. The strong and consistent association between continuity of care and GPs' perceptions of the suitability of VCs is a new and important finding with considerable relevance for future primary health care planning. In accordance with existing literature and guidance, GPs' perceived suitability of VCs varied considerably across reasons for contact and presented health problems. The findings cannot be directly generalized beyond the specific context of the pandemic lockdown, but nevertheless provide interesting results regarding the performance of VCs for different reasons for contact and clinical conditions. The results indicate that GPs still consider the physical examination a crucial element of many consultations to enhance both diagnostic accuracy and quality of care in a wider sense. Reflecting on the accumulated experience with VCs, most participating GPs envisaged conducting 20% of their consultations by video in a future, nonpandemic setting.

Acknowledgments

After the survey was conducted, TMJ and BLN held part-time positions at the Norwegian Centre for E-health Research in Tromsø, Norway. FLO 2.0 (a GP-driven, nonprofit organization dedicated to the sustainable development of Norwegian general practice) covered the license to Netigate.

Norwegian Health Informatics made a crucial contribution with the dispatch of the survey through their subscriber email list and website. The Norwegian Directorate of eHealth and Confrere provided background data on video consultations.

Authors' Contributions

TMJ and NLB developed the survey together with LG, FHK, and BA. EK and PZ subsequently joined the team and contributed substantially to data analysis, interpretation of results, and writing, together with the rest of the authors.

Conflicts of Interest

The authors declare no conflicts of interests. The commercial actors who distributed the survey (Norwegian Health Informatics) and contributed information on the use of VCs in Norway in spring 2020 (Confrere) were at no point involved in the design of the survey or data analysis.

Multimedia Appendix 1
Consultations with general practitioners in Norway during the COVID-19 lockdown. The shaded period corresponds to the period of the data collection. April 6-12, 2020, was the Easter holiday week. Data obtained from Norwegian general practitioners' reimbursement claims and The Norwegian Directorate of eHealth.
[PNG File , 53 KB - jmir_v23i2e26433_app1.png]

Multimedia Appendix 2
Digital consultations with general practitioners in Norway during the COVID-19 lockdown. The shaded period corresponds to the period of data collection. April 6-12, 2020, was the Easter holiday week. Data from Norwegian general practitioners' reimbursement claims and The Norwegian Directorate of eHealth.
[PNG File , 154 KB - jmir_v23i2e26433_app2.png]

References

1. Digital First Primary Care. NHS England. URL: https://www.england.nhs.uk/gp/digital-first-primary-care/ [accessed 2021-01-25]
2. Scotland's digital health and care strategy: enabling, connecting and empowering. Scottish Government. 2018. URL: https://www.gov.scot/publications/scotlands-digital-health-care-strategy-enabling-connecting-empowering/ [accessed 2021-01-25]
3. Helsedirektoratet, Direktoratet for e-helse. E-konsultasjon hos fastlege og legevakt [E-consultations with the GP and the emergency room] 2018. URL: https://www.helsedirektoratet.no/rapporter/e-konsultasjon-helsedirektoratet-og-direktoratet-for-e-helse [accessed 2021-01-25]
4. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. The potential of alternatives to face-to-face consultation in general practice, and the impact on different patient groups: a mixed-methods case study. Health Serv Deliv Res 2018 Jun;6(20):1-200. [doi: 10.3310/hsdr06200] [Medline: 29889485]
5. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for COVID-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
6. Webster P. Virtual health care in the era of COVID-19. The Lancet 2020 Apr;395(10231):1180-1181. [doi: 10.1016/s0140-6736(20)30818-7]
7. Mold F, Hendy J, Lai Y, de Lusignan S. Electronic Consultation in Primary Care Between Providers and Patients: Systematic Review. JMIR Med Inform 2019 Dec 03;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]
8. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020 Mar 17;4(1):bjgpopen20X101020. [doi: 10.3399/bjgpopen20x101020]
9. Atherton H, Brant H, Ziebland S, Bikker A, Campbell J, Gibson A, et al. Alternatives to the face-to-face consultation in general practice: focused ethnographic case study. Br J Gen Pract 2018 Jan 29;68(669):e293-e300. [doi: 10.3399/bjgp18x694853]
10. Mueller M, Knop M, Niehaves B, Adarkwah CC. Investigating the Acceptance of Video Consultation by Patients in Rural Primary Care: Empirical Comparison of Preusers and Actual Users. JMIR Med Inform 2020 Oct 22;8(10):e20813 [FREE Full text] [doi: 10.2196/20813] [Medline: 32969339]
11. Randhawa RS, Chandan JS, Thomas T, Singh S. An exploration of the attitudes and views of general practitioners on the use of video consultations in a primary healthcare setting: a qualitative pilot study. Prim Health Care Res Dev 2018 Jun 18;20:1. [doi: 10.1017/s1463423618000361]
12. Powell RE, Henstenburg JM, Cooper G, Hollander JE, Rising KL. Patient Perceptions of Telehealth Primary Care Video Visits. Ann Fam Med 2017 May 08;15(3):225-229 [FREE Full text] [doi: 10.1370/afm.2095] [Medline: 28483887]
13. Greenhalgh T, Morrison C, Choon G. Video consultations: a guide for practice 2020. URL: https://bjgplife.com/2020/03/18/video-consultations-guide-for-practice/ [accessed 2021-01-25]
14. Saultz JW. Defining and measuring interpersonal continuity of care. Ann Fam Med 2003 Sep 01;1(3):134-143 [FREE Full text] [doi: 10.1370/afm.23] [Medline: 15043374]

15. Pereira Gray DJ, Sidaway-Lee K, White E, Thorne A, Evans PH. Continuity of care with doctors-a matter of life and death? A systematic review of continuity of care and mortality. BMJ Open 2018 Jun 28;8(6):e021161 [FREE Full text] [doi: 10.1136/bmjopen-2017-021161] [Medline: 29959146]
16. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q 2005 Sep;83(3):457-502 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00409.x] [Medline: 16202000]
17. Gray DP, Freeman G, Johns C, Roland M. Covid 19: a fork in the road for general practice. BMJ 2020 Sep 28;370:m3709. [doi: 10.1136/bmj.m3709] [Medline: 32988832]
18. Helsedirektoratet, Direktoratet for e-helse. E-konsultasjon og triage hos fastlege [E-consultations and triage with the GP]. 2019. URL: https://www.helsedirektoratet.no/rapporter/e-konsultasjon-helsedirektoratet-og-direktoratet-for-e-helse [accessed 2021-01-25]
19. Helse-og omsorgsdepartementet. Oppfordrer helsetjenesten til å bruke videokonsultasjon [Encourages the health care service to use video consultation] 2020. URL: https://www.regjeringen.no/no/aktuelt/oppfordrer-helsetjenesten-til-a-bruke-videokonsultasjon/id2693883/ [accessed 2021-01-25]
20. Tikkanen R, Osborn R, Mossialos E, Djordjevic A, Wharton G. The Commonwealth Fund. International Health Care System Profils - Norway 2020. URL: https://www.commonwealthfund.org/international-health-policy-center/countries/norway [accessed 2021-01-25]
21. Car J, Koh GC, Foong PS, Wang CJ. Video consultations in primary and specialist care during the COVID-19 pandemic and beyond. BMJ 2020 Oct 20;371:m3945. [doi: 10.1136/bmj.m3945] [Medline: 33082127]
22. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
23. Norsk Helseinformatikk AS. URL: https://nhi.no/ [accessed 2021-01-25]
24. Helsedirektoratet, Fastlegestatistikk [Statistics for Norwegian General Practitioners]. URL: https://www.helsedirektoratet.no/statistikk/fastlegestatistikk [accessed 2021-01-27]
25. Salisbury C, Quigley A, Hex N, Aznar C. Private Video Consultation Services and the Future of Primary Care. J Med Internet Res 2020 Oct 01;22(10):e19415 [FREE Full text] [doi: 10.2196/19415] [Medline: 32812887]
26. Donaghy E, Atherton H, Hammersley V, McNeilly H, Bikker A, Robbins L, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Jun 03;69(686):e586-e594. [doi: 10.3399/bjgp19x704141]
27. Haun MW, Stephan I, Wensing M, Hartmann M, Hoffmann M, Friederich H. Intent to Adopt Video-Based Integrated Mental Health Care and the Characteristics of its Supporters: Mixed Methods Study Among General Practitioners Applying Diffusion of Innovations Theory. JMIR Ment Health 2020 Oct 15;7(10):e23660 [FREE Full text] [doi: 10.2196/23660] [Medline: 33055058]
28. Johnsen T, Norberg B, Krogh F, Sigurdsson J, Getz L. Complex issues in general practice - a prevalence study. Tidsskr Nor Laegeforen 2020 Jun 30;140(10):1017-1022. [doi: 10.4045/tidsskr.19.0683] [Medline: 32602308]
29. Andrees V, Klein T, Augustin M, Otten M. Live interactive teledermatology compared to in-person care - a systematic review. J Eur Acad Dermatol Venereol 2020 Apr;34(4):733-745. [doi: 10.1111/jdv.16070] [Medline: 31715035]
30. Brinker TJ, Hekler A, von Kalle C, Schadendorf D, Esser S, Berking C, et al. Teledermatology: Comparison of Store-and-Forward Versus Live Interactive Video Conferencing. J Med Internet Res 2018 Oct 24;20(10):e11871 [FREE Full text] [doi: 10.2196/11871] [Medline: 30355564]
31. Chuchu N, Dinnes J, Takwoingi Y, Matin R, Bayliss S, Davenport C, Cochrane Skin Cancer Diagnostic Test Accuracy Group. Teledermatology for diagnosing skin cancer in adults. Cochrane Database Syst Rev 2018 Dec 04;12(12):CD013193 [FREE Full text] [doi: 10.1002/14651858.CD013193] [Medline: 30521686]
32. Malterud K, Reventlow S, Guassora AD. Diagnostic knowing in general practice: interpretative action and reflexivity. Scand J Prim Health Care 2019 Dec 11;37(4):393-401 [FREE Full text] [doi: 10.1080/02813432.2019.1663592] [Medline: 31507239]
33. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-World Implementation of Video Outpatient Consultations at Macro, Meso, and Micro Levels: Mixed-Method Study. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
34. Marshall M, Howe A, Howsam G, Mulholland M, Leach J. COVID-19: a danger and an opportunity for the future of general practice. Br J Gen Pract 2020 May 11;70(695):270-271. [doi: 10.3399/bjgp20x709937]
35. Salisbury C, Murphy M, Duncan P. The Impact of Digital-First Consultations on Workload in General Practice: Modeling Study. J Med Internet Res 2020 Jun 16;22(6):e18203 [FREE Full text] [doi: 10.2196/18203] [Medline: 32543441]
36. Newbould J, Abel G, Ball S, Corbett J, Elliott M, Exley J, et al. Evaluation of telephone first approach to demand management in English general practice: observational study. BMJ 2017 Sep 27;358:j4197 [FREE Full text] [doi: 10.1136/bmj.j4197] [Medline: 28954741]
37. Sinha S, Kern LM, Gingras LF, Reshetnyak E, Tung J, Pelzman F, et al. Implementation of Video Visits During COVID-19: Lessons Learned From a Primary Care Practice in New York City. Front Public Health 2020 Sep 17;8:514 [FREE Full text] [doi: 10.3389/fpubh.2020.00514] [Medline: 33042950]

38. Hammersley V, Donaghy E, Parker R, McNeilly H, Atherton H, Bikker A, et al. Comparing the content and quality of video, telephone, and face-to-face consultations: a non-randomised, quasi-experimental, exploratory study in UK primary care. Br J Gen Pract 2019 Jul 01;69(686):e595-e604. [doi: 10.3399/bjgp19x704573]
39. Shaw SE, Seuren LM, Wherton J, Cameron D, A'Court C, Vijayaraghavan S, et al. Video Consultations Between Patients and Clinicians in Diabetes, Cancer, and Heart Failure Services: Linguistic Ethnographic Study of Video-Mediated Interaction. J Med Internet Res 2020 May 11;22(5):e18378 [FREE Full text] [doi: 10.2196/18378] [Medline: 32391799]

Abbreviations

GP: general practitioner
VC: video consultation

Edited by G Eysenbach; submitted 11.12.20; peer-reviewed by E Opoku-Agyemang; comments to author 06.01.21; revised version received 13.01.21; accepted 18.01.21; published 08.02.21.

Please cite as:
Johnsen TM, Norberg BL, Kristiansen E, Zanaboni P, Austad B, Krogh FH, Getz L
Suitability of Video Consultations During the COVID-19 Pandemic Lockdown: Cross-sectional Survey Among Norwegian General Practitioners
J Med Internet Res 2021;23(2):e26433
URL: http://www.jmir.org/2021/2/e26433/
doi:10.2196/26433
PMID:33465037

Original Paper

Technology, Privacy, and User Opinions of COVID-19 Mobile Apps for Contact Tracing: Systematic Search and Content Analysis

Mahmoud Elkhodr[1], PhD; Omar Mubin[2], PhD; Zainab Iftikhar[3], BS; Maleeha Masood[3], BS; Belal Alsinglawi[2], MSc; Suleman Shahid[3], PhD; Fady Alnajjar[4], PhD

[1]School of Engineering and Technology, Central Queensland University, Sydney, Australia

[2]School of Computer, Data and Mathematical Sciences, Western Sydney University, Rydalmere, Australia

[3]Department of Computer Science, Syed Babar Ali School of Science and Engineering, Lahore University of Management Sciences, Lahore, Pakistan

[4]Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, Alain, United Arab Emirates

Corresponding Author:
Fady Alnajjar, PhD
Department of Computer Science and Software Engineering
College of Information Technology
United Arab Emirates University
Alain 15551
Alain,
United Arab Emirates
Phone: 971 037135538
Email: fady.alnajjar@uaeu.ac.ae

Abstract

Background: Many countries across the globe have released their own COVID-19 contact tracing apps. This has resulted in the proliferation of several apps that used a variety of technologies. With the absence of a standardized approach used by the authorities, policy makers, and developers, many of these apps were unique. Therefore, they varied by function and the underlying technology used for contact tracing and infection reporting.

Objective: The goal of this study was to analyze most of the COVID-19 contact tracing apps in use today. Beyond investigating the privacy features, design, and implications of these apps, this research examined the underlying technologies used in contact tracing apps. It also attempted to provide some insights into their level of penetration and to gauge their public reception. This research also investigated the data collection, reporting, retention, and destruction procedures used by each of the apps under review.

Methods: This research study evaluated 13 apps corresponding to 10 countries based on the underlying technology used. The inclusion criteria ensured that most COVID-19-declared epicenters (ie, countries) were included in the sample, such as Italy. The evaluated apps also included countries that did relatively well in controlling the outbreak of COVID-19, such as Singapore. Informational and unofficial contact tracing apps were excluded from this study. A total of 30,000 reviews corresponding to the 13 apps were scraped from app store webpages and analyzed.

Results: This study identified seven distinct technologies used by COVID-19 tracing apps and 13 distinct apps. The United States was reported to have released the most contact tracing apps, followed by Italy. Bluetooth was the most frequently used underlying technology, employed by seven apps, whereas three apps used GPS. The Norwegian, Singaporean, Georgian, and New Zealand apps were among those that collected the most personal information from users, whereas some apps, such as the Swiss app and the Italian (Immuni) app, did not collect any user information. The observed minimum amount of time implemented for most of the apps with regard to data destruction was 14 days, while the Georgian app retained records for 3 years. No significant battery drainage issue was reported for most of the apps. Interestingly, only about 2% of the reviewers expressed concerns about their privacy across all apps. The number and frequency of technical issues reported on the Apple App Store were significantly more than those reported on Google Play; the highest was with the New Zealand app, with 27% of the reviewers reporting technical difficulties (ie, 10% out of 27% scraped reviews reported that the app did not work). The Norwegian, Swiss, and US (PathCheck) apps had the least reported technical issues, sitting at just below 10%. In terms of usability, many apps, such as those from Singapore, Australia, and Switzerland, did not provide the users with an option to sign out from their apps.

Conclusions: This article highlighted the fact that COVID-19 contact tracing apps are still facing many obstacles toward their widespread and public acceptance. The main challenges are related to the technical, usability, and privacy issues or to the requirements reported by some users.

(J Med Internet Res 2021;23(2):e23467) doi:10.2196/23467

KEYWORDS

contact tracing; COVID-19; digital contact tracing apps

Introduction

Overview

The COVID-19 pandemic, the virus of which causes a highly contagious respiratory infection, has spread rapidly across the world and surpassed 20 million cases by early August 2020 [1]. The economic impact of the pandemic is felt globally with many countries slipping into recession. The COVID-19 pandemic is also turning into a job crisis, which is threatening to dismantle several industries, from aviation and manufacturing to services, tourism, and agriculture [2].

The global public health and government responses to the pandemic have been fragmented due to the urgency of actions required as a result of the stochastic spread of the virus. Some countries are implementing policies to eradicate the virus, such as Vietnam and New Zealand [3]; some countries are trying to suppress and contain the spread of the virus, such as Australia [4]; and some countries are relying on building herd immunity, such as Sweden [5]. Nonetheless, the virus continues to spread arbitrarily between regions and countries, and the epicenter of the pandemic has been moving between continents. It started with China and moved to Italy, Spain, the United States, and Brazil, with India as the next in line. Several other countries are now experiencing a second wave after initially suppressing it, with clusters of new cases popping up in many countries [6].

The speed of the authorities' response has also proven to be a major key in containing the spread of the virus. For instance, many experts weighed in on the relatively slow response of Italy to contain the virus [7] and the fast response of South Korea in suppressing it [8]. Despite the variations in the worldwide governmental crisis responses to the pandemic and the lack of clear and uniform advice on matters as simple as the role of a mask in containing the spread of the virus [9], the measures and policies used worldwide to contain the virus remained mostly precautionary in the absence of a vaccine or a treatment. Consequently, the direct safety advice as a result of the COVID-19 pandemic continues to be about maintaining good hand hygiene, practicing social distancing between people, testing as soon as virus symptoms appear, quarantining, and, importantly, contact tracing.

Contact tracing is the process of identifying, assessing, and managing people who have been exposed to a disease to prevent onward transmission [10]. Until a COVID-19 vaccine is commercially available to the public, contact tracing tools are vital in breaking the chains of transmission of the virus. This means identifying infected people and their close contacts, testing them, and isolating them for 14 days from day zero of the exposure. For countries that managed to control the exponential growth of the virus, known as flattening the curve, extensive contact tracing was essential in minimizing large-scale community transmissions. With countries recently coming out of lockdown and opening their economies and borders again, such as France and the United States, contact tracing is the key to rapidly identifying new cases; hence, maintaining low levels of community transmissions to remain successful in containing the outbreak of the virus. Thus, in addition to comprehensive testing capacity, contact tracing is increasingly becoming important in managing this pandemic until a vaccine or a reliable viral treatment is successful and made publicly available.

For contact tracing to be beneficial in preventing onward transmission, and thereby reducing the impact of a second wave of a contagious disease such as COVID-19, it should be implemented systematically. This means having a system to securely collect, compile, and analyze data about individuals in real time, while not impinging on their privacy. As with the lack of a uniform and standardized global response to the pandemic, contact tracing technologies and approaches adopted by several countries were also diverse. For instance, on the same day in which Canada announced that they were working on a new contact tracing app [11], the United Kingdom was abandoning their contact tracing app, stating that the technology does not work [12].

Background

Contact tracing using a mobile app relies on the concept of proximity tracking. The concept behind contact tracing is to identify and keep a record of people who may have been in close proximity (eg, typically less than 1.5 meters) to other people. Therefore, once an individual is identified to be infected with COVID-19, the app will be used to retrieve and trace the other close contacts. There have been various implementations for contact tracing apps, and a range of technologies, security, and privacy approaches have been adopted across the globe. Notably, the effectiveness of these contact tracing technologies remains to be seen. More evidence is required to demonstrate whether these tools were successful in contact tracing and to determine their usefulness.

Before the COVID-19 pandemic, a range of digital and mobile health tools had been utilized for the purposes of infectious disease control and public health interventions. Aba et al [13] illustrated the variety of functionalities provided by mobile apps to mitigate the spread of the Ebola virus in Africa, ranging from contract tracing to surveillance to case management. Mobile apps have also had a similar range of success in the use of public health interventions mitigating the spread of tuberculosis in Botswana [14].

The current contact tracing apps for COVID-19, which have been widely used by several countries, mostly use Bluetooth as the underlying technology for proximity sensing. In an effort to contribute toward having a unified solution for contact tracing and to counter the limitations of using Bluetooth on the iOS platform [15], Apple and Google have also recently released a new framework to support contact tracing [16]. However, apps that implement this framework have not matured enough yet. Nonetheless, surveying the current apps in use; analyzing their privacy features, penetration, and intake; and measuring their reception by the public, including the ensuing issues faced, have not been fully explored. This is demonstrated in a survey of the prior literature, which is presented in Table S1 from Multimedia Appendix 1. We briefly summarize a review of the main studies from the literature below.

The user acceptability of contact tracing apps in five countries hit by the pandemic using a survey were investigated in Altmann et al [17]; however, the study did not review specific contact tracing apps. Similarly, several studies [18-22] did not review current COVID-19 apps through direct access of app stores. For instance, the work reported in Anglemyer et al [18] did a meta-analysis on medical databases to review contact tracing apps. Others used various methodologies to conduct their reviews.

Other works, such as the one reported in Collado-Borrell at al [23], attempted to identify smartphone apps that aimed to address the COVID-19 pandemic and analyzed their characteristics. However, the study did not investigate any specific app. It only classified the apps under specific categories, such as health, fitness, or medicine. The main security and data protection aspects relating to digital contact tracing frameworks and apps were also investigated in Martin at al [24]; the paper analyzed some of the privacy aspects, such as personal information access, data retention, and location tracking. The paper also highlighted some of the app's public penetration; however, the study was only limited to apps from Google Play.

An overview of mobile apps being currently used for COVID-19 and their assessment using the Mobile Application Rating Scale was reported in Davalbhakta et al [25]. This study was limited only to India, the United States, and the United Kingdom. Other works, such as the one reported in Vaudena [26], studied Bluetooth-based contact tracing solutions, including Decentralized Privacy-Preserving Proximity Tracing (DP-3T) and Temporary Contact Numbers (TCN) protocols, from a centralized versus decentralized point of view. As such, the vulnerabilities and the advantages of both solutions were systematically reviewed. The work focused more on the underlying architecture used and the level of privacy protection each one presented; however, it did not review specific implementations of COVID-19 apps. Only a few apps were used to represent the centralized and decentralized approaches.

The work in Magklaras and Bojorquez [22] surveyed the data regulations and technology protocols relating to COVID-19 contact tracing apps. It also provided mapping for the global deployment of the COVID-19 contact tracing apps. The paper also discussed the challenges, including some privacy aspects, relating to Bluetooth-based contact tracking technologies. The work reported in Li and Guo [27] provided an in-depth review of COVID-19 tracing app technologies and processes, including app installation and registrations, encounter data processing and communication, and notifications. The paper also analyzed the security aspects of contact tracing app architectures (ie, centralized, decentralized, and hybrid) by assessing their risk against common security attacks, such as denial of service and carryover attacks. This paper as well as an additional review paper [28] discussed some users' common concerns, but it did not qualitatively analyze any users' reviews.

Study Aims

To this end, the work presented in this paper reviews and evaluates most categories of COVID-19 contact tracing mobile apps in use today. To our knowledge, this is the first research study that primarily investigated the public's and users' perceptions of COVID-19 contact tracing apps. This study also aimed at studying the privacy feature implementations and the level of penetration these apps achieved. In extension to the first aim, we aimed to determine the outreach of the collated apps in terms of number of downloads, as reported not only by the app stores but also by the authorities of each of the apps' corresponding countries. This is in addition to providing a quantitative overview of the common complaints suggested by app users in connection to privacy, battery drainage, technical difficulties, bugs, crashes, and more. Additionally, in relation to the second aim, the underlying apps' architecture and associated aspects, such as how the communication or handshake between two devices in proximity took place and then how close contacts were reported, were also analyzed. We also investigated the timeline of when these apps were introduced. Lastly, extending from the third aim, we attempted to understand the nature, type, and extent of data capture of the apps, such as granularity of data that was captured (ie, location, identification, and accomplices), duration of data retention, option to discard and delete records, and whether opt-out options were provided to the user without uninstalling the app.

Methods

Selection of Apps, User Intake, and Penetration

This study classified contact tracing apps based on the type of technology used for contact tracing of infected masses. This study identified six distinct technologies and an additional category commonly used or incorporated into COVID-19 tracing apps. These included Bluetooth, the DP-3T protocol, GPS, Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT), the TCN protocol, Google and Apple, and other technologies, mainly the use of Quick Response (QR) codes paired with a digital diary. These technologies are outlined in Table S2 from Multimedia Appendix 1.

The classification criteria considered the underlying technology used by the apps rather than classifying the apps based on geographical or other architectural features. This is because most of the apps in use today use Bluetooth. Therefore, classifying the apps based on the underlying technology ensures that the research is capturing most contact tracing solutions in use. For instance, contact tracing solutions used by Singapore, Australia, and Malaysia use the same technology (ie, BlueTrace).

As such, there is little benefit to the research from surveying all three of these apps.

Therefore, the research evaluated 13 apps corresponding to 10 countries and covered all the contact tracing technologies identified above. All apps were free to download. The inclusion criteria also ensured that most of the COVID-19-declared epicenters (ie, countries) were included in the sample, such as Italy and the United States. The evaluated apps also included countries that did relatively well in controlling the outbreak of COVID-19, such as Singapore; countries that had a low daily number of new infections (ie, Australia); and countries that had a medium-level daily number of new infections (ie, Pakistan). The Swiss app was included in this study, as Switzerland was among the few countries that did not implement a lockdown. Similarly, the Swedish app was also included, given Sweden's unique approach to building herd immunity to combat COVID-19. Informational apps or unofficial contact tracing apps were excluded from this study, except for South Korea's Corona 100m app, which uses GPS technology for contact tracing. This app was included because Corona 100m was among the first major contact tracing apps that launched across the globe and because South Korea is one of the few countries that managed to quickly suppress the transmission of the virus. Figure 1 shows the apps that were included in this study.

Figure 1. The 13 apps corresponding to 10 countries included in this study. DP-3T: Decentralized Privacy-Preserving Proximity Tracing; PEPP-PT: Pan-European Privacy-Preserving Proximity Tracing; QR: Quick Response.

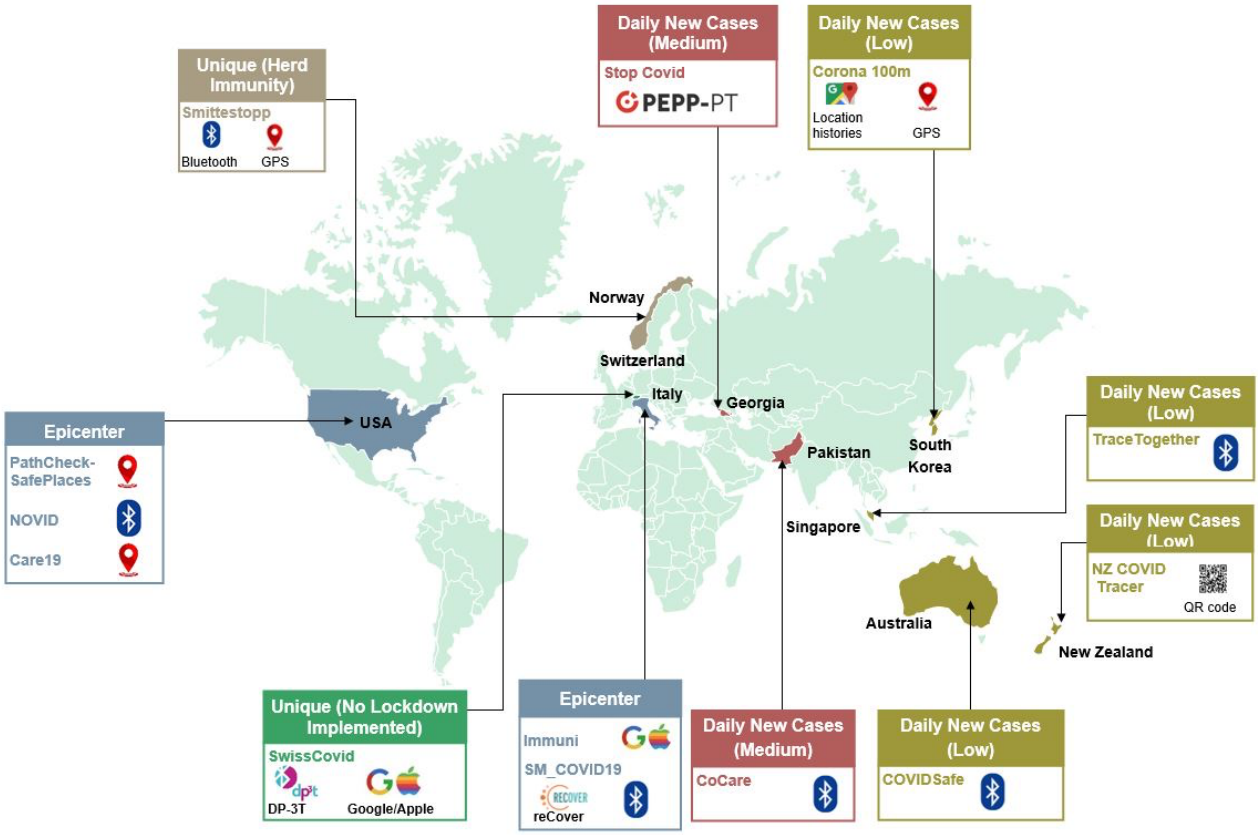


Table 1 details the architecture and approaches used by each of these technologies; these are as follows:

1. Country: For each type of technology used, a sample of countries that use this technology and their contact tracing apps are provided. Where there is more than one app used in a country, the name of the corresponding app is provided. It is worth noting that this is not a comprehensive list. The aim is to evaluate some of the countries for the purpose of adding context to the data presented in the table rather than creating an inventory of apps. The next section provides more details on the selection and inclusion criteria of the apps evaluated in this study.
2. Architecture: This criterion investigates whether or not the technology used by the contact tracing app incorporates the concept of uploading contact logs to a central reporting server. The criteria used are *centralized*, *semicentralized*, and *decentralized*. It has proven difficult to exclusively classify the architecture of each of these technologies, as implementations varied from one app to another. For instance, some apps uploaded contact logs to a central server, but the server did not have access to the uploaded contact logs, nor was it responsible for any further contact tracing processing, while others had access. As such, this criterion should be read in conjunction with the other criteria presented in Table 1, mainly the *encounter handshake* and *infection reporting* criteria.
3. Encounter handshake: This refers to how two devices coming into close contact perform a handshake (ie, exchange identification data). Most of the technologies surveyed exchanged some form of a temporary ID, while others exchanged some form of a unique identifier that was either encrypted or in plain text, which also depended on the specific implementation of each of the apps.

4. Infection reporting. This refers to how the contact log is reported to the central server and the role of this server in contact tracing. Most of the apps relied on the users to upload the contact logs. Implementations varied as well based on whether the health authorities had access to the contact logs or not.
5. Privacy by design. As the name suggests, this criterion explored whether the technology embedded any privacy considerations into its design specifications.

To analyze the users' intake of each of the 13 apps under review and the penetration by these apps, this study extracted the following data for each of the apps: the name and country where the app was launched, the number of installs as per Google Play and as reported by the local news in the home country of the app, the penetration percentage as per Google Play installs and as reported by local news sources, and the launch date of each of the apps.

The number of installs were not only sourced from Google Play but also from local news outlets from the home country of each of the corresponding apps. The penetration percentages sourced from Google Play and the ones extracted from local news sources were calculated by dividing the total number of installs by the total population of the home country.

Table 1. The technologies of the contact tracing apps and their salient features.

App information	App technology						
	Bluetooth	DP-3T[a]	GPS	PEPP-PT[b]	TCN[c]	Google and Apple	Others
Example countries[d]	Australia; Singapore; Malaysia (MyTrace)	Austria; Finland; The Netherlands	Iceland (Ranking C-19); Italy (Diary); Jordan (Aman)	France; Georgia; Italy (Immuni)	Germany (ITO); Italy; The United States	Canada; Switzerland (SwissCovid); Germany (Corona-Warn-App)	New Zealand (digital diary); Australia, Canada, and New Zealand (GetHomeSafe); Malaysia (SELangkah)
Architecture	Centralized	Decentralized	Centralized	Centralized	Semicentralized	Decentralized	Centralized
Encounter handshake	Users exchange temporary IDs issued by the server	Unique 128-bit pseudorandom identifier by the server	Varies by implementation; some identify users by phone numbers	Users exchange temporary IDs issued by the server	TCN	Unique identifiers that are encrypted with a secret daily key held by the sending device	Real ID
Infection reporting	User-triggered upload	User-triggered upload, but the health authority never has access to contact log	User-triggered upload	User-triggered upload, but received massive privacy backlashes	The app notifies the user to potential infection	Not provisioned; delegated to app implementation	Varies by implementation; mostly user triggered
Privacy by design	No	Yes	No	No	Yes	Yes	No

[a]DP-3T: Decentralized Privacy-Preserving Proximity Tracing.

[b]PEPP-PT: Pan-European Privacy-Preserving Proximity Tracing.

[c]TCN: Temporary Contact Numbers.

[d]Where there was more than one app used in a country, the name of the corresponding app is provided within parentheses.

Investigating the Privacy-by-Design Features and Privacy Implementations of COVID-19 Contact Tracing Apps

This study expands on previous work [29] that compared the privacy aspects of the COVIDSafe app (Australia, Bluetooth) and the COVID Tracer app (New Zealand, QR code). Each of the selected apps was downloaded and evaluated thoroughly. The study first identified the underlying technology used for contact tracing by an app and the amount of personal information each app collected (ie, personal information access). To do that, the following scale was used: if an app was only collecting the name, email, and phone number of the user, then the scale was designated as *low*; if, in addition to this personal information, the app collected the age of the user, then the scale was designated as *medium*; and if an app collected the name, email, phone number, age, and any additional information, such as the address, ethnicity, or location via GPS of the user, then this criterion was rated as *high*.

Additionally, the study analyzed the location features and tracking capabilities for each of the apps. It investigated whether an app was tracking the movement of individuals or not (ie, location tracking). It also investigated whether the app under review knew the identity of the people in close proximity to the user or just their locations or IDs (ie, true identity vs temporary ID, such as with the TCN protocol). The criterion *tracking and identifying proxies* combined the *encounter handshake* and *infection reporting* features.

Furthermore, this study investigated the record-keeping time frame of each app. This was achieved by researching the duration that the contact logs were kept on the device or the authority's remote servers for each of the apps under review.

In terms of user control, this study examined two criteria: the user's right to forget and the geo-restrictions imposed on an app. The first criterion considered whether or not users were informed about the procedures to delete the records collected by an app. The opting-out criterion explored whether the users were able to sing in and out of the app under review. Lastly, for each app under review, the study investigated whether an app could be downloaded from anywhere or whether it was a home or region geo-restricted app (ie, geo-restriction). We referred to governments' media releases, white papers, and developers' announcements for the apps that were in testing phases or were not available on the Apple App Store or Google Play.

Analyzing the Public Reception of COVID-19 Contact Tracing Apps

We aimed to identify the audience uptake and users' feedback of the COVID-19 contact tracing apps under review. Data were sourced by scraping the publicly available user reviews from the Apple App Store and Google Play webpages for each of the apps. Scraping is a process or tool used to extract data from a website; in this case, reviews from Google and Apple stores. Almost 30,000 reviews were scraped and analyzed in this study. The user reviews of each of the corresponding apps were then filtered and analyzed using a brute-force keyword search methodology; this means extracting the user reviews that contained a specific keyword used in the search. Table 2 lists the keywords used in scraping the reviews. The methodology used for analyzing these reviews also accounted for the variations of each of the keywords, referred to as subkeywords. For instance, the results of scraping and analyzing certain subkeywords—*doesn't work*, *didn't work*, *not working*, *Doesn't work*, *Didn't work*, and *Not working*—were all counted toward the results of the main keyword *Malfunctioning*. In other words, the results reported under the keyword *Malfunctioning* are a concatenation of each of the individual results returned by its list of subkeywords.

Table 2. The keywords used in this study.

Keywords	Subkeywords
Drainage	drain; battery; Drain; Battery
Spyware	spy; spied; spyware; Spy; Spied; Spyware
Malfunctioning	doesn't work; didn't work; not working; Doesn't work; Didn't work; Not working
Crashes	crash; freeze; Crash; Freeze
Privacy concerns	privacy issue; privacy concern; location concern; tracking me; track me; tracking us; Privacy issue; Privacy concern; Location concern; Tracking me; Track me; Tracking us
Ineffective	useless; rubbish; garbage; Useless; Rubbish; Garbage
Bugs	bug; buggy; Bug; Buggy
Installation issues	can't install; doesn't install; couldn't install; Can't install; Doesn't install; Couldn't install
Incompatible	can't download; couldn't download; incompatible; Can't download; Couldn't download; Incompatible

Results

Selection of Apps, User Intake, and Penetration

In this section, we initially describe results on app penetration. A challenging aspect of sourcing the data reported in Table 3 [30-40] was encountered when calculating the intake of the apps under study. For instance, the number of downloads for an app does not represent the true value of the actual intake. Downloading an app does not necessarily mean the app is being used. Users may simply download the app and never use it or uninstall it. In addition, there were little data available on the number of uninstalls for each of the surveyed apps. Regardless of this limitation, the number of installations for an app was not available on the Apple App Store. This has made the task of calculating the uptake of an app even more complex.

Consequently, the research required access to a more precise estimate of the installation values as compared to what Google Play was showing. Therefore, apart from consulting Google Play's number of installs, the study referred to reliable news sources to obtain the total number of registrations or downloads for each of the apps under review. The news sources were mainly from government or developer announcements, verifiable local news sources, and published research (ie, white papers). Some of the statistical information, such as the download intakes and any data sourced from local news, was available as of early July 2020. As such, there might be a slight variation in the values presented in Table 3 as compared those at the time of the archiving of this paper. Some apps were new, so this local value was not readily available for those either. Another challenge this research study ran into was the unavailability of some of the apps on the Google Play Store. This is because they were discontinued or because they were still in demo or beta stages.

Table 3. Penetration and intake of the 13 selected contact tracing apps.

Country	App	No. of installs, n		Penetration, %		No. of days of the app's launch since patient zero[a], n
		Local news	Google Play Store[b]	Local news	Google Play Store[b]	
United States	PathCheck SafePlaces	N/A[c]	10,000	N/A	0.001	93
United States	NOVID	N/A	10,000	N/A	0.001	110
United States	Care19	33,000 [31]	10,000	0.01	0.001	76
Italy	Immuni	2,700,000 [32]	1,000,000	4.47	1.65	122
Italy	SM-COVID-19	52,000 [33]	50,000	0.09	0.08	73
Norway	Smittestopp	1,427,000 [34]	100,000	26.32	1.84	50
Singapore	TraceTogether	2,100,000 [35]	1,000,000	35.89	17.09	57
South Korea	Corona 100m	1,000,000 [36]	N/A	1.95	N/A	20
Pakistan	CoCare	N/A	500	N/A	0.001	108
Australia	COVIDSafe	6,130,000 [37]	1,000,000	24.03	3.92	91
New Zealand	NZ COVID Tracer	573,000 [38]	100,000	11.88	2.07	82
Switzerland	SwissCovid	1,600,000 [39]	500,000	18.48	5.78	90
Georgia	Stop Covid	100,000 [40]	100,000	2.51	N/A	50

[a]The first case in the country was reported on the Johns Hopkins Coronavirus Resource Center portal [30].

[b]All download values have been extracted from the app's webpage on Google Play.

[c]N/A: not applicable; data for apps were unavailable because the app was new, in the case of local news, or because it was discontinued or was still in demo or beta stages, in the case of the Google Play Store.

The results of this study show that South Korea was the first country to use a mobile app for contact tracing during the COVID-19 pandemic. The South Korean app, Corona 100m, was only introduced after 20 days from the first detected case of COVID-19 in South Korea. This was followed by Singapore, Norway, and Georgia, which introduced their apps around 50 days since patient zero. The United States, Italy, and Pakistan were slower, as they introduced their contact tracing app around the 100-day mark. As reported by local news, the Singaporean (36%) and Norwegian (26%) apps had the highest penetration intake, followed by the Australian and Swiss apps, which had around 20% penetration, and the New Zealand app, which achieved around 11% penetration. Interestingly, the Italian and US apps had the lowest penetration values. The penetration intake of the apps on the Android platform, which was calculated based on the Google Play–reported number of installs for each of the apps, showed that all apps under review, except for the Singaporean app, had very poor intake (<5%).

Furthermore, this research initially intended to calculate the success rate of each of the apps in contact tracing reporting. It also aimed to survey and compare the efficacy of the apps under review. However, this was challenged by the lack of any reliable relevant data available in relation to those aspects; thus, this part of the review had to be excluded. Therefore, it is unclear whether the early introduction of contact tracing apps has contributed toward their rapid public adoption or whether these apps have played a major role in the contact tracing efforts of COVID-19. Perhaps these apps have played some part in raising awareness among the public, as in the case of the Singaporean and Australian apps, which had higher penetration intake values and lower infection numbers comparatively. However, due to insufficient data, no conclusive results can be made on the correlation between the early introduction of a contact tracing app, its higher penetration intake, and the case where low number of COVID-19 transmissions were reported.

Investigating the Privacy-by-Design Features and Privacy Implementations of COVID-19 Contact Tracing Apps

In the subsequent sections in this paper, when referring to an app, the following notation shall be used: app name (country of origin, technology used for contact tracing).

Table S3 from Multimedia Appendix 1 reviews the privacy features of the 13 apps evaluated in this study. Each of these apps was downloaded and evaluated thoroughly as per the criteria shown in Table S3 from Multimedia Appendix 1. The research also referred to white papers and developers' announcements for the apps that were in their testing phases or were not available or accessible on the Apple App Store and/or Google Play. The same methodology was followed for the apps that were not available in English, such as Immuni (Italy, Google and Apple application programming interface [API]) [41], SM-COVID-19 (Italy, ReCoVer) [42], and Smittestopp (Norway, Bluetooth and GPS) [43].

Nine of the apps were available for free on both the Apple App Store and Google Play. Two apps—SM-COVID-19 (Italy, Google and Apple) and CoCare (Pakistan, Bluetooth) [44]—were only available on Google Play, while Stop Covid (Georgia, PEPP-PT) [45] was only available on the Apple App

Store. The Corona 100m app (South Korea, location) [46] was not available on both stores. Smittestopp (Norway, Bluetooth and GPS) was not available to download due to geo-restrictions. The Australian COVIDSafe app required an Australian phone number and a postcode to run.

Bluetooth was the most frequently used underlying technology, employed by seven apps for digital contact tracing, whereas three apps performed contact tracing through location (eg, GPS). The apps using location as the underlying technology, namely Corona 100m (South Korea, location) and PathCheck SafePlaces (United States, location), tracked and recorded the locations visited by the users. Although Corona 100m (South Korea, location) was removed from Google Play, the app integrated GPS history, data from nationwide surveillance cameras, and credit card transactions. This has sparked privacy concerns, as users of the Corona 100m app could see the date when a COVID-19 patient was infected, along with his or her nationality, gender, age, and the locations they visited.

The Norwegian, Singaporean, Georgian [45], and New Zealand [47] apps were among the apps that collected the most personal information from the users, while some other apps, such as the Swiss app [48] and the Italian Immuni app, did not collect any user information. Other apps ranged from simply collecting users' phone numbers to additionally collecting their names or email addresses.

Data destruction was incorporated into most of the apps, which automatically deleted the users' records after 14 days, the observed minimum amount of time implemented in most of the apps. Some kept these records for 21 days (ie, Australia) and others for 30 days (ie, Switzerland and India); the New Zealand app kept them for 31 days, while the Georgian apps kept them for 3 years, the longest of any app.

Three of the US apps—PathCheck (United States, location) [49], NOVID (United States, Bluetooth radio waves and ultrasound) [50], and Care19 (United States, GPS) [51]—did not require users to sign up before using their app. On the other hand, many apps, such as the Singaporean TraceTogether app [52], the Australian COVIDSafe app [53], and the Swiss and Indian apps, did not provide the users with an option to sign out from their app. It is noteworthy to mention that the data presented in Table S3 of Multimedia Appendix 1 are accurate as of June 30, 2020.

Analyzing the Public Reception of COVID-19 Contact Tracing Apps

Figure 2 shows the percent occurrence for each of the keywords for each app. Figure 3 shows the average ratings of the reviews for each keyword. For example, consider if a user left a review for one of the apps saying, "the app keeps on crashing," and then gave it a rating of 2 stars. This review will then be counted toward the average mentions of the keyword *crashes* shown in Figure 2. The 2-star rating will also be counted toward the corresponding keyword average rating shown in Figure 3. All small values were rounded up to 0.001.

Figure 2. Percent occurrence of each keyword for each app. NA: not applicable; user reviews were unavailable, as the corresponding apps were not available on the corresponding platforms.

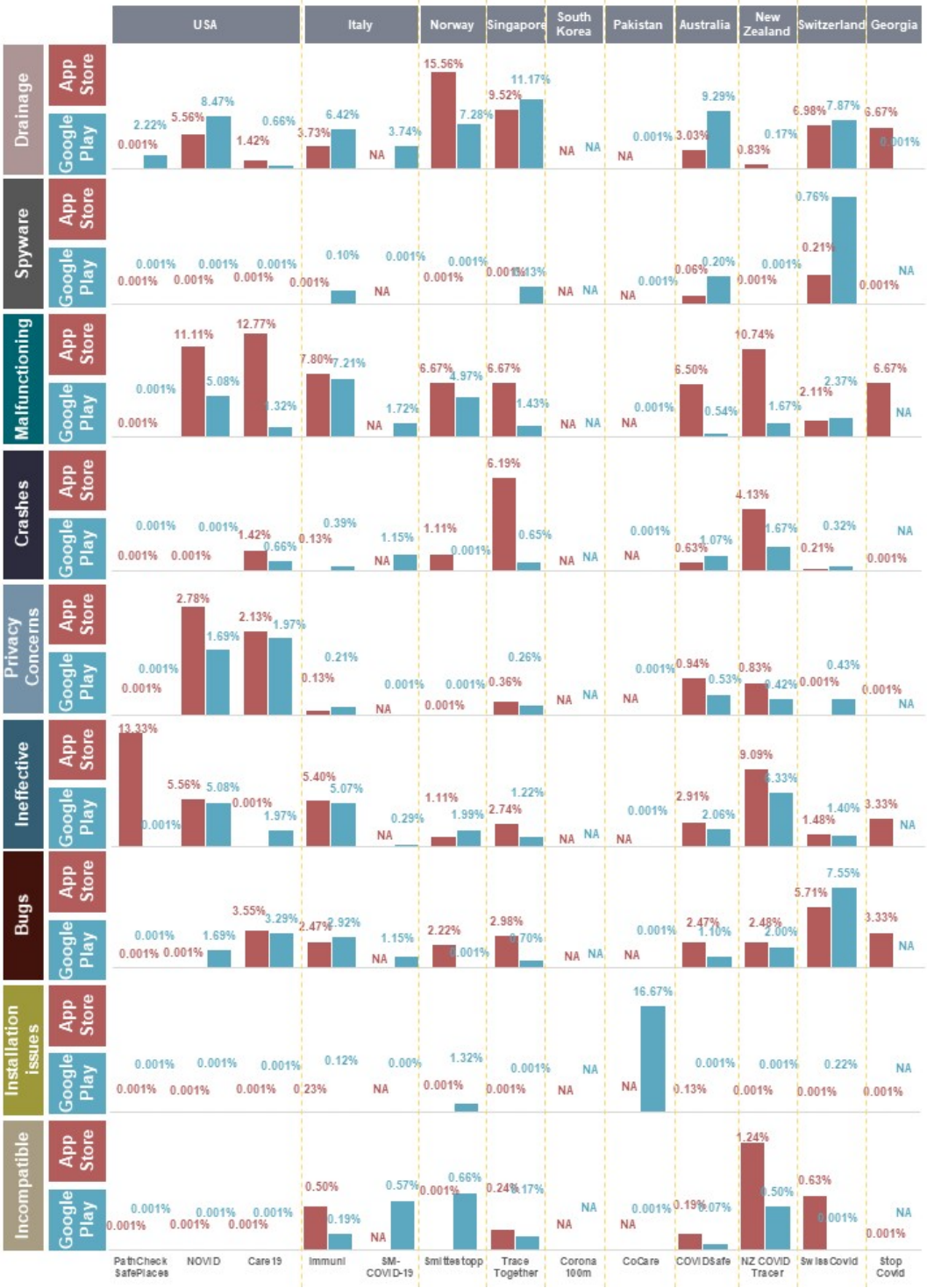
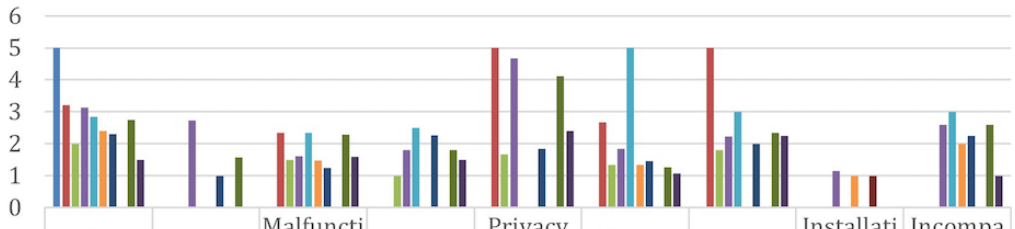


Figure 3. Average ratings out of 5 stars from user reviews in each category of each app on Google Play.



	Drainage	Spyware	Malfunctioning	Crashes	Privacy Concerns	Effective	Bugs	Installation	Incompatible
PathCheck SafePlaces	5	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
NOVID	3.2	0.001	2.333	0.001	5	2.667	5	0.001	0.001
Care19	2	0.001	1.5	1	1.667	1.333	1.8	0.001	0.001
Immuni	3.127	2.727	1.61	1.81	4.682	1.845	2.231	1.154	2.6
SM-COVID-19	2.846	0.001	2.333	2.5	0.001	5	3	0.001	3
Smittestopp	2.409	0.001	1.467	0.001	0.001	1.333	0.001	1	2
Trace Together	2.311	1	1.242	2.267	1.833	1.464	2	0.001	2.25
CoCare	0.001	0.001	0.001	0.001	0.001	0.001	0.001	1	0.001
COVIDSafe	2.751	1.571	2.289	1.8	4.108	1.257	2.338	0.001	2.6
NZ COVID Tracer	1.5	0.001	1.6	1.5	2.4	1.079	2.25	0.001	1

Three of the applications—CoCare (Pakistan, Bluetooth), SM-COVID-19 (Italy, ReCoVer), and Corona 100m (South Korea, location)—were not available on the Apple App Store, whereas two apps—the Corona 100m (South Korea, location) and Stop Covid (Georgia, PEPP-PT)—were not available on Google Play. Based on the frequency of keyword occurrences, *Drain*, *Malfunctioning*, and *Ineffective* were the most frequent issues reported by the users in their reviews.

On the Apple App Store, the keyword *rubbish* had a 13.33% occurrence for PathCheck SafePlaces (United States, location), 5.56% for NOVID (United States, Bluetooth), 5.40% for Immuni (Italy, Google and Apple API), and 9.09% for NZ COVID Tracer (New Zealand, digital diary). Similarly, many users did not find contact tracing apps functional. On the Apple App Store, many app users complained that their app did not work. This was represented by the keyword *Malfunctioning*, which had a 10.74% occurrence for NZ COVID Tracer (New Zealand, digital diary), 6.50% for COVIDSafe (Australia, Bluetooth), 6.67% for TraceTogether (Singapore, Bluetooth), 7.80% for Immuni (Italy, Google and Apple API), 11.11% for NOVID (United States, Bluetooth), and a sharp 12.77% occurrence for Care19 (United States, Apple and Google). Many users also had problems with the apps' compatibility with their operating system and frequent crashes. For instance, CoCare (Pakistan, Bluetooth) had a 16.67% occurrence for the incompatibility issue.

Interestingly, and as shown in Figure 4, no significant battery drainage issue had been reported for most of the reviewed apps. The privacy concerns reported by the users were also very minimal across all apps, as shown in Figure 5.

Figure 4. Percent occurrence of the keyword "drainage," pertaining to battery drainage.

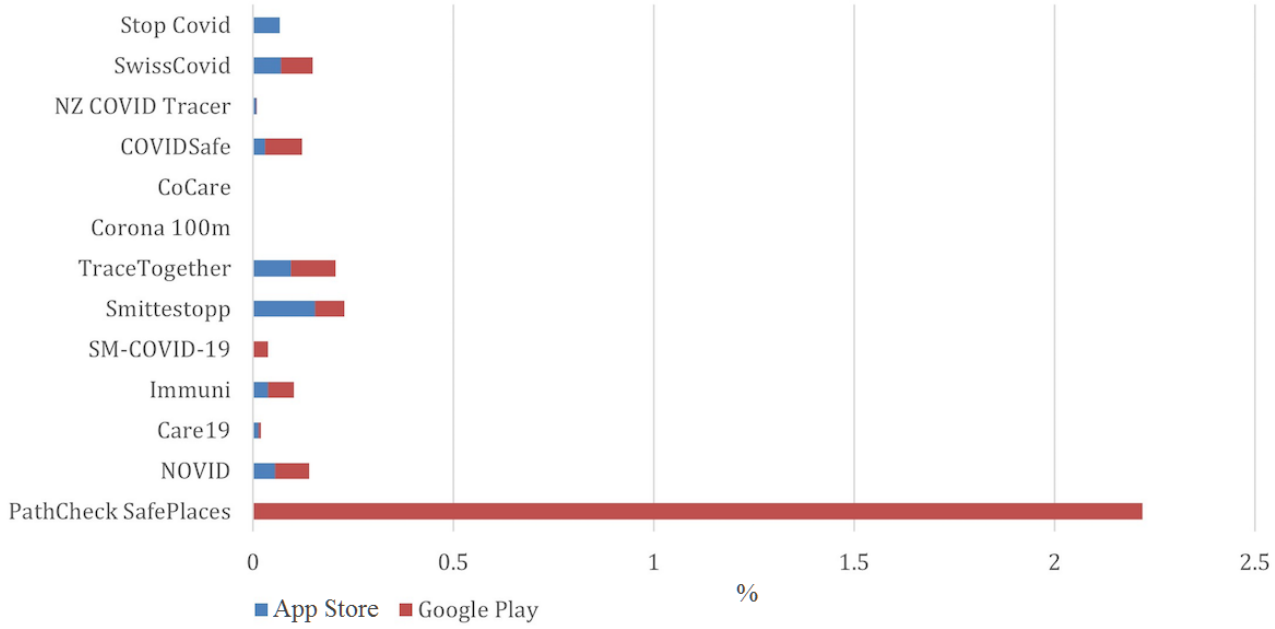
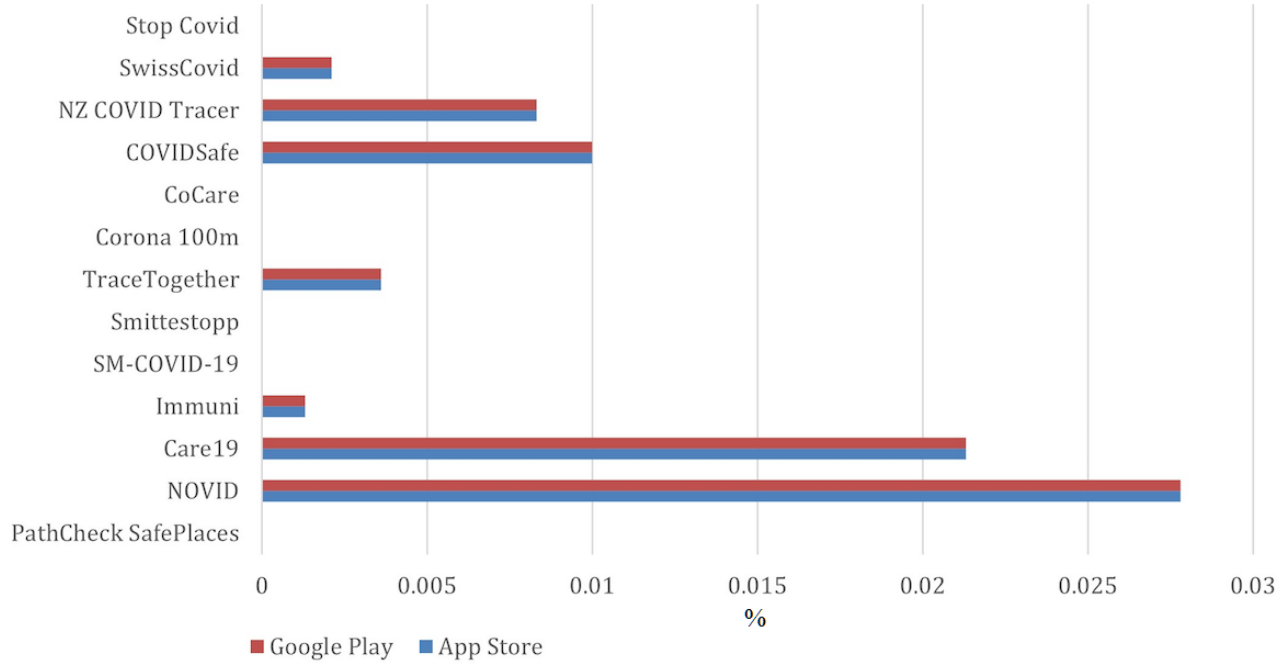


Figure 5. Percent occurrence of the keywords "privacy concerns" and "spyware".



Figures 6 and 7 provide overall insights into the technical issues reported by the users for each of the apps. These figures combine the results of the following keywords, along with their respective subkeywords: *Malfunctioning, Crashes, Ineffective, Bugs, Installation issues, and Incompatible*. It is obvious that most apps on the Apple App Store had the most reported technical issues when compared to their Google Play counterparts, except for the Swiss contact tracing app. The US PathCheck app had the least reported technical issues on Google Play, while the New Zealand app version on the Apple App Store had the most technical issues that were complained about across all apps and platforms.

Figure 6. Summary of technical issues reported for each of the apps. The plot shows results from the combination of the following keywords, along with their respective subkeywords: "Malfunctioning," "Crashes," "Ineffective," "Bugs," "Installation issues," and "Incompatible".

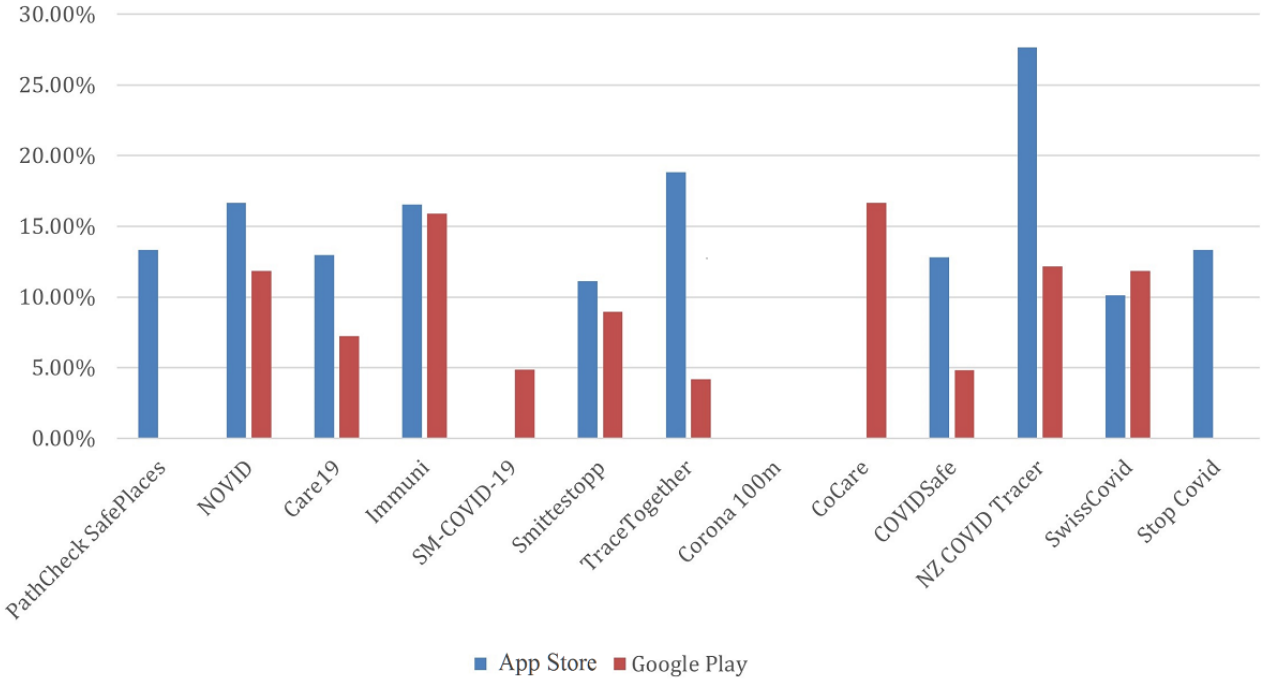
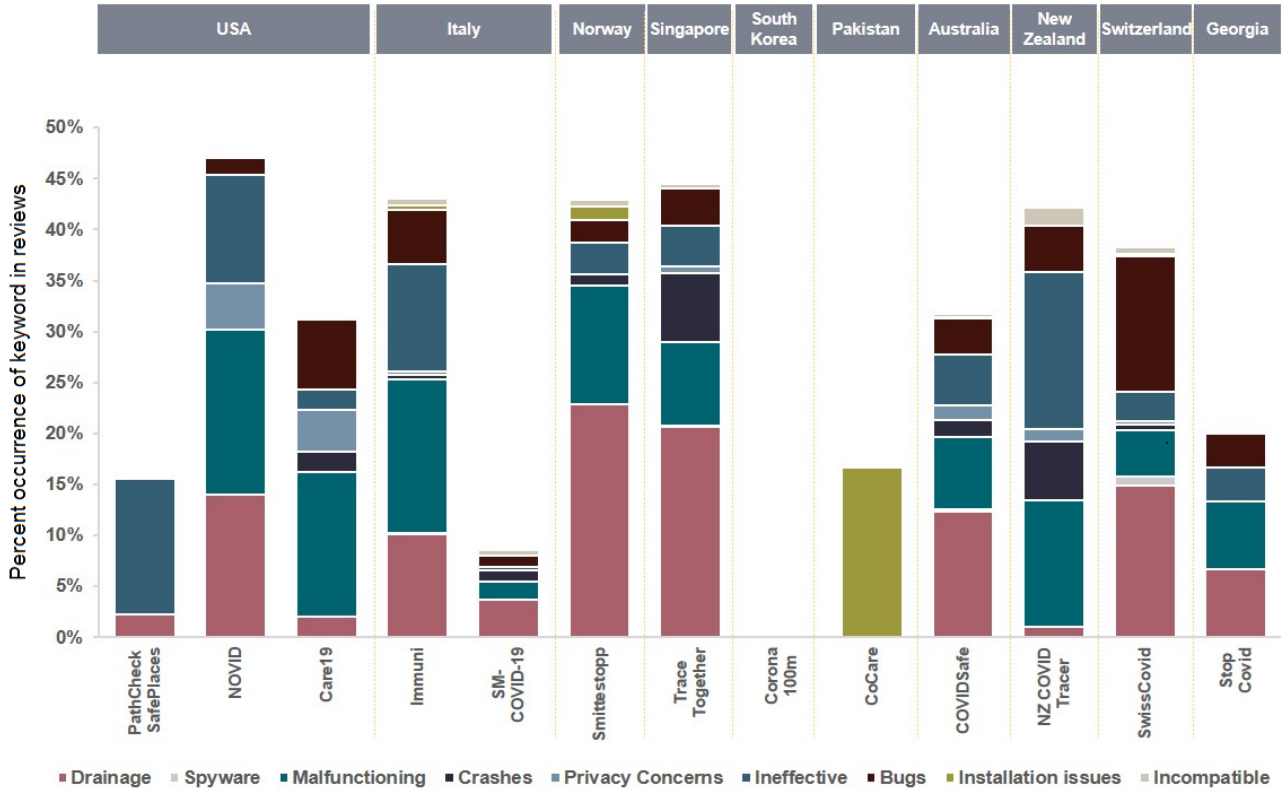


Figure 7. Comparison of user app reviews and their inclusion of various keywords.



Discussion

Principal Findings

Our research study has highlighted the hindrances in the successful deployment of COVID-19 contact tracing apps. The use of mobile technologies for contact tracing has been met with a number of challenges [54], many of which also emerged in the contextual evaluation of user reviews on COVID-19 contact tracing apps as described in our study. Among these, the most popular were technical malfunctions and drainage of battery.

Other challenges included privacy. Of course, this is anticipated, as you cannot expect to trace and track peoples' movements by a government authority without addressing privacy issues [55].

Nonetheless, in addition to privacy, there were many other challenges and limitations hindering the anticipated efficacy from contact tracing apps.

For instance, a mobile contact tracing app needs to be widely adopted by a population for it to be of benefit; this is challenging to achieve. Penetration of COVID-19 contact tracing apps remains low despite governments pushing for mass use [56]. The widespread adoption of contact tracing apps requires that people would have access to a smartphone and, in most cases, access to a reliable internet connection. Hence, in countries with large populations, such as Pakistan [57], the smartphone penetration percentage sits at only 15%, and in Indonesia this value sits at only 31%. Users may also feel uncomfortable if there is no clear opt-out strategy [58]. Nearly half of the apps reviewed in our study did not provide a transparent withdrawal avenue.

Furthermore, the approaches used by contact tracing apps rely mostly on one single parameter (ie, proximity, such as via Bluetooth) [59]. However, proximity by itself is not enough to determine the risk of someone being exposed to the virus. There are a number of other parameters involved, such as being indoors or outdoors, being in a room with good air circulation or not, and the issue of surface infection exposure, irrespective of the proximity of an individual to an infected person. Furthermore, although as shown in our review that Bluetooth was one of the more popular technologies to implement contact tracing apps, every country's regulations may differ [60]; hence, a one-size-fits-all approach may be problematic.

Other challenges pertain to the limitations associated with the technology used for contact tracing. For instance, the use of GPS as a proximity technology is not reliable in indoor environments [61]. Determining the distance between two persons using Bluetooth technology also has its own set of challenges, such as signal strength attenuation caused by some environmental factors (eg, if the phone is placed inside a thick pocket or if the phone is at an angle facing a wall).

Nevertheless, contact tracing technologies surveyed in this work have been found to use a locationless tracking approach; that is, the app does not trace or record people's movements, obviously for privacy purposes. Therefore, most of these apps can only determine if two people were in proximity at a given time, but they do not keep a log of the users' movements. Consider, for example, if an infected person, labelled as Pi, is in a supermarket and Pi touches an item at time t–1 at a location designated as Li. Another person who is not infected, designated as Pn, is at a location designated as Ln. There is no proximity between Pi and Pn. Now assume Pi leaves the store at time t, when at the same time (ie, at t) person Pn moves from Ln to Li. There is a high chance that Pn is going to be infected if they touch the same item Pi touched at t–1 (ie, surface infection exposure). To be able to capture this exposure, contact tracing apps require the use of a location-oriented tracking approach in which the locations and movements of people are compared against each other to determine the overlapped and colluded locations. Future work will explore the use of our already well-established location obfuscation technique [62] in a contact tracing solution. The work will aim at providing a location-oriented contact tracing app without impinging on users' privacy.

Limitations

One of the challenges encountered in scraping the reviews was analyzing the apps that were not available in English. For example, most of the reviews for the Immuni app (Italy, Google and Apple), SM-COVID-19 (Italy, ReCoVer), and Smittestopp (Norway, Bluetooth and GPS) were available in the Italian and Norwegian languages, respectively. For these reviews, along with the rest of the app reviews that were in different languages, the keywords along with their subkeywords were translated into the language of their home app country. The results were incorporated when calculating the overall average values for all the apps. The translated keywords along with the subkeywords used can be found in Table S4 from Multimedia Appendix 1. Another limitation in our methodology for review scraping lies in the presence of false negatives in some of the reviews. This is one of the limitations of brute-force keyword search methodology. Take, for instance, one of the reviews for COVIDSafe (Australia, Bluetooth) on Google Play:

> *Installed from its release. Worked. No problems at all. It doesn't drain the battery. It doesn't crash. It's totally fine. I haven't been dragged into the back of a van, taken to an underground bunker and questioned by spies.*

The review is classified as a false negative for the words *drain* and *crash*. It can be debated that the number of false negatives could have been reduced by simply taking the *battery* subkeyword out from the keyword search (ie, battery; drain). However, in doing so, the number of 1-star reviews were significantly reduced by more than 50%. For instance, with NZ COVID Tracer (New Zealand, digital diary), the 1-star reviews dropped from 23 to 10 after taking the word *battery* out of the search filter. The reason behind this is that the users' reviews were not systematic. Most users represented their opinions in natural language. Some samples of 1-star reviews for COVIDSafe (Australia, Bluetooth) commenting on the app's drainage issue are as follows:

> *It is of no use whatsoever. A waste of money & a waste of my battery life.*
>
> *Battery went from 100% to zero in 5 hours with not much use. I usually get a full day out of it.*
>
> *Hard on the battery.*

Therefore, for the sake of including these comments, the subkeyword *battery* was not removed from the keyword search results.

This study has a number of additional limitations. This paper is based solely on 13 evaluated apps. While the selection criteria ensured that apps were selected to represent each of the categories of technology used for contact tracing, it did not review all COVID-19 contact tracing apps. Also, the study relied on data that were extracted and accurate as of July 2020. Another major limitation of this work relates to the penetration intake calculations done for each of the apps. The study derived the percent penetration for each app by dividing the number of installs of an app by the population of the home country. This

method suffers from several shortcomings. The number of downloads or installs cannot be precise; it ignores the fact that some users may install, uninstall, and reinstall the app several times or on more than one device owned by the user. A user may also download the app and never use it. Many users may not download an app yet may still post a review about it. For these reasons, the study attempted to gauge the penetration intake for each of the apps by analyzing the local government announcements and reports published by local news agencies. However, since the reported download values cannot be verified nor the methods used to derive them, the trustworthiness of these values also remains invalidated. Lastly, although the study evaluated about 30,000 user reviews, the reviews cannot not be verified.

Conclusions

While public health agencies attempt to understand the efficacy of nonpharmaceutical interventions [63], contact tracing has been a key part of the worldwide measure in response to the COVID-19 pandemic. For contact tracing to work effectively, solutions such as tracing apps should be implemented systematically. This requires the secure collection, processing, storage, and discarding of contact tracing information of people in real time, without impinging on their privacy and rights. The success of contact tracing apps greatly depends on their large uptake within a population, in addition to strong public health enforcement. This article highlighted the fact that COVID-19 contact tracing apps are still facing many obstacles toward their widespread and public acceptance.

The main challenges are related to the technical, usability, and privacy issues or requirements reported by some users. This meant that most tracing apps were not publicly well-received and had low penetration levels, which hinders their effectiveness. For instance, only the Singaporean app had a penetration of slightly over 30%, the Australian and Swiss apps achieved penetration just below 20%, and the penetration values reported for most of the other apps were very poor, sitting at below 5%. The amount of personal data collected by the apps varied widely, with some apps not collecting data at all and others collecting a significant amount of sensitive data about the user, such as their ethnicity. The majority of the surveyed apps did not provide the user with options to opt out from the apps, such as logging out, without uninstalling them.

The lack of a standardized contact tracing approach also meant that contact tracing apps used across the globe were fragmented and noninteroperable. As most countries are now coming out of lockdown and reopening their borders, there is an increased need for a cohesive, cross-border, and interoperable contact tracing app that can be used universally without impinging on users' privacy. Additionally, there is a lack of available data on the effectiveness of COVID-19 contact tracing apps. As we progressively recuperate from this pandemic, there is a need to re-evaluate and re-examine the values and roles of contact tracing apps in controlling infectious diseases such as COVID-19.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of papers related to contact tracing apps, the main technologies used, the privacy features, and translated keywords.
[DOCX File , 47 KB - jmir_v23i2e23467_app1.docx]

References

1. COVID-19 situation update worldwide. European Centre for Disease Prevention and Control. 2020. URL: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases [accessed 2020-08-11]
2. Akkermans J, Richardson J, Kraimer ML. The Covid-19 crisis as a career shock: Implications for careers and vocational behavior. J Vocat Behav 2020 Jun;119:103434 [FREE Full text] [doi: 10.1016/j.jvb.2020.103434] [Medline: 32390655]
3. Baker MG, Kvalsvig A, Verrall AJ. New Zealand's COVID-19 elimination strategy. Med J Aust 2020 Sep;213(5):198-200.e1 [FREE Full text] [doi: 10.5694/mja2.50735] [Medline: 32789868]
4. Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. Nat Commun 2020 Nov 11;11(1):5710 [FREE Full text] [doi: 10.1038/s41467-020-19393-6] [Medline: 33177507]
5. Orlowski EJW, Goldsmith DJA. Four months into the COVID-19 pandemic, Sweden's prized is nowhere in sight. J R Soc Med 2020 Aug;113(8):292-298. [doi: 10.1177/0141076820945282] [Medline: 32780968]
6. Strzelecki A. The second worldwide wave of interest in coronavirus since the COVID-19 outbreaks in South Korea, Italy and Iran: A Google Trends study. Brain Behav Immun 2020 Aug;88:950-951 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.042] [Medline: 32311493]
7. Ren X. Pandemic and lockdown: A territorial approach to COVID-19 in China, Italy and the United States. Eurasian Geogr Econ 2020 May 08;61(4-5):423-434 [FREE Full text] [doi: 10.1080/15387216.2020.1762103]
8. Reis RF, de Melo Quintela B, de Oliveira Campos J, Gomes JM, Rocha BM, Lobosco M, et al. Characterization of the COVID-19 pandemic and the impact of uncertainties, mitigation strategies, and underreporting of cases in South Korea,

Italy, and Brazil. Chaos Solitons Fractals 2020 Jul;136:109888 [FREE Full text] [doi: 10.1016/j.chaos.2020.109888] [Medline: 32412556]
9. Chiu NC, Chi H, Tai YL, Peng CC, Tseng CY, Chen CC, et al. Impact of Wearing Masks, Hand Hygiene, and Social Distancing on Influenza, Enterovirus, and All-Cause Pneumonia During the Coronavirus Pandemic: Retrospective National Epidemiological Surveillance Study. J Med Internet Res 2020;22(8):e21257 [FREE Full text] [doi: 10.2196/21257] [Medline: 32750008]
10. Du L, Raposo VL, Wang M. COVID-19 Contact Tracing Apps: A Technologic Tower of Babel and the Gap for International Pandemic Control. JMIR Mhealth Uhealth 2020;8(11):e23194 [FREE Full text] [doi: 10.2196/23194] [Medline: 33156804]
11. Kleinman RA, Merkel C. Digital contact tracing for COVID-19. CMAJ 2020 Jun 15;192(24):E653-E656 [FREE Full text] [doi: 10.1503/cmaj.200922] [Medline: 32461324]
12. Sabbagh D, Hern A. UK abandons contact-tracing app for Apple and Google model. The Guardian. 2020 Jun 18. URL: https://www.theguardian.com/world/2020/jun/18/uk-poised-to-abandon-coronavirus-app-in-favour-of-apple-and-google-models [accessed 2021-01-28]
13. Tom-Aba D, Nguku PM, Arinze CC, Krause G. Assessing the concepts and designs of 58 mobile apps for the management of the 2014-2015 West Africa Ebola outbreak: Systematic review. JMIR Public Health Surveill 2018 Oct 29;4(4):e68 [FREE Full text] [doi: 10.2196/publichealth.9015] [Medline: 30373727]
14. Littman-Quinn R, Chandra A, Schwartz A, Fadlelmola F, Ghose S, Luberti A, et al. mHealth applications for telemedicine and public health intervention in Botswana. In: Proceedings of the IST-Africa Conference. 2011 Presented at: IST-Africa Conference; May 11-13, 2011; Gaborone, Botswana p. 1-11 URL: https://ieeexplore.ieee.org/document/6107376
15. Cunche M, Boutet A, Castelluccia C, Lauradoux C, Roca V. On Using Bluetooth-Low-Energy for Contact Tracing. Montbonnot-Saint-Martin, France: Inria Grenoble Rhône-Alpes; INSA de Lyon; 2020. URL: https://hal.inria.fr/hal-02878346v5/document [accessed 2021-01-29]
16. Michael K, Abbas R. Behind COVID-19 contact trace apps: The Google–Apple partnership. IEEE Consum Electron Mag 2020 Sep 1;9(5):71-76. [doi: 10.1109/mce.2020.3002492]
17. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: Cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
18. Anglemyer A, Moore T, Parker L, Chambers T, Grady A, Chiu K. Digital contact tracing technologies in epidemics: A rapid review. Cochrane Database Syst Rev 2020:1-46. [doi: 10.1002/14651858.cd013699]
19. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
20. Kondylakis H, Katehakis DG, Kouroubali A, Logothetidis F, Triantafyllidis A, Kalamaras I, et al. COVID-19 Mobile Apps: A Systematic Review of the Literature. J Med Internet Res 2020;22(12):e23170 [FREE Full text] [doi: 10.2196/23170] [Medline: 33197234]
21. Baraniuk C. Covid-19 contact tracing: A briefing. BMJ 2020 May 13;369:m1859. [doi: 10.1136/bmj.m1859] [Medline: 32404304]
22. Magklaras G, Bojorquez L. A review of information security aspects of the emerging COVID-19 contact tracing mobile phone applications. In: Proceedings of the International Symposium on Human Aspects of Information Security and Assurance. 2020 Presented at: International Symposium on Human Aspects of Information Security and Assurance; July 7-9, 2021; Mytilene, Greece p. 30-44. [doi: 10.1007/978-3-030-57404-8_3]
23. Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M. Features and functionalities of smartphone apps related to COVID-19: Systematic search in app stores and content analysis. J Med Internet Res 2020 Aug 25;22(8):e20334 [FREE Full text] [doi: 10.2196/20334] [Medline: 32614777]
24. Martin T, Karopoulos G, Hernández-Ramos JL, Kambourakis G, Nai Fovino I. Demystifying COVID-19 digital contact tracing: A survey on frameworks and mobile apps. Wirel Commun Mob Comput 2020 Oct 17;2020:1-29. [doi: 10.1155/2020/8851429]
25. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A systematic review of smartphone applications available for coronavirus disease 2019 (COVID19) and the assessment of their quality using the Mobile Application Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
26. Vaudenay S. Centralized or decentralized? The contact tracing dilemma. OpenAIRE. 2020 Jun 01. URL: https://covid-19.openaire.eu/search/publication?articleId=od_____185::b333a9cdf196cdc396fede6971196acc [accessed 2021-01-29]
27. LI J, Guo X. Global deployment mappings and challenges of contact-tracing apps for COVID-19. SSRN 2020 May 24:1-7 [FREE Full text] [doi: 10.2139/ssrn.3609516]
28. Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, et al. A survey of COVID-19 contact tracing apps. IEEE Access 2020;8:134577-134601. [doi: 10.1109/access.2020.3010226]
29. Elkhodr M. New Zealand's COVID-19 Tracer app won't help open a 'travel bubble' with Australia anytime soon. The Conversation. 2020 May 20. URL: https://theconversation.com/

new-zealands-covid-19-tracer-app-wont-help-open-a-travel-bubble-with-australia-anytime-soon-139026 [accessed 2021-01-28]
30. COVID-10 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JSU). Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2021-01-28]
31. Howell O'Neill P. Why one US state will have two coronavirus tracing apps. MIT Technology Review. 2020 May 20. URL: https://www.technologyreview.com/2020/05/20/1002042/why-one-us-state-will-have-two-coronavirus-tracing-apps/ [accessed 2021-01-28]
32. D'Alessandro J. La app anti pandemia Immuni vicina ai tre milioni di download. la Repubblica. 2020 Jun 16. URL: https://www.repubblica.it/tecnologia/2020/06/16/news/immuni_vicina_ai_tre_milioni_di_download-259372693/ [accessed 2021-01-28]
33. SM-COVID-19. URL: https://smcovid19.org/ [accessed 2021-01-28]
34. Nikel D. Norway: 1.4 million people download coronavirus tracking app despite security concerns. Forbes. 2020 Apr 25. URL: https://www.forbes.com/sites/davidnikel/2020/04/25/norway-14-million-people-download-coronavirus-tracking-app-despite-security-concerns/?sh=fb00dc57832c [accessed 2021-01-28]
35. TraceTogether. Government of Singapore. URL: https://www.tracetogether.gov.sg/ [accessed 2021-01-28]
36. Dujmovic J. Wildly popular coronavirus-tracker app helps South Koreans steer clear of outbreak areas. MarketWatch. 2020 Mar 18. URL: https://www.marketwatch.com/story/wildly-popular-coronavirus-tracker-app-helps-south-koreans-steer-clear-of-outbreak-areas-2020-03-18 [accessed 2021-01-28]
37. Meixner S. How many people have downloaded the COVIDSafe app and how central has it been to Australia's coronavirus response? ABC News. 2020 Jun 01. URL: https://www.abc.net.au/news/2020-06-02/coronavirus-covid19-covidsafe-app-how-many-downloads-greg-hunt/12295130 [accessed 2021-01-28]
38. Changelog: Covid Tracing Tracker updates as they happen. MIT Technology Review. 2020 May 15. URL: https://www.technologyreview.com/2020/05/15/1001736/changelog-covid-tracing-tracker-updates-as-they-happen/ [accessed 2021-01-28]
39. More than 1.6 million Swiss have downloaded Covid tracing app. The Local CH. 2020 Jun 29. URL: https://www.thelocal.ch/20200629/more-than-800000-downloads-for-switzerlands-covid-app-in-three-days [accessed 2021-01-28]
40. More than 100 thousand people downloaded covonavirus tracking app in Georgia. Georgian Journal. 2020 Apr 27. URL: https://www.georgianjournal.ge/society/36537-more-than-100-thousand-people-downloaded-covonavirus-tracking-app-in-georgia.html [accessed 2021-01-28]
41. Immuni, Ministero della Salute. Google Play. URL: https://play.google.com/store/apps/details?id=it.ministerodellasalute.immuni&hl=en_AU&gl=US [accessed 2021-01-28]
42. SM_Covid19, Softmining Srl. Google Play. URL: https://play.google.com/store/apps/details?id=it.softmining.projects.covid19.savelifestyle&hl=en_AU&gl=US [accessed 2021-01-28]
43. Smittestopp, Folkehelseinstituttet. Google Play. URL: https://play.google.com/store/apps/details?id=no.fhi.smittestopp_exposure_notification&hl=en_AU&gl=US [accessed 2021-01-28]
44. CoCare, Cheetay. Google Play. URL: https://play.google.com/store/apps/details?id=com.cocareapp&hl=en_AU&gl=US [accessed 2021-01-28]
45. Stop Covid - let's fight this together, Ministry of IDPs, Labour, Health, Social Affairs. Google Play. URL: https://webcache.googleusercontent.com/search?q=cache:12nc9hpdEu8J:https://play.google.com/store/apps/details%3Fid%3Dgov.georgia.novid20%26hl%3Den_CA%26gl%3DUS+&cd=1&hl=en&ct=clnk&gl=ca [accessed 2021-01-28]
46. Review of mobile application technology to enhance contact tracing capacity for COVID-19. Johns Hopkins Center for Health Security. 2020 Apr 08. URL: https://www.centerforhealthsecurity.org/resources/COVID-19/COVID-19-fact-sheets/200408-contact-tracing-factsheet.pdf [accessed 2021-01-28]
47. NZ COVID Tracer app. Ministry of Health NZ. 2020. URL: https://www.health.govt.nz/our-work/diseases-and-conditions/covid-19-novel-coronavirus/covid-19-resources-and-tools/nz-covid-tracer-app [accessed 2021-01-28]
48. SwissCovid, Federal Office of Public Health FOPH. Google Play. URL: https://play.google.com/store/apps/details?id=ch.admin.bag.dp3t&hl=en_AU&gl=US [accessed 2021-01-28]
49. SafePlaces, Path Check, Inc. Google Play. URL: https://play.google.com/store/apps/details?id=org.pathcheck.covidsafepaths&hl=en_AU&gl=US [accessed 2021-01-28]
50. NOVID, Expii, Inc. Google Play. URL: https://play.google.com/store/apps/details?id=com.expii.novid&hl=en_AU&gl=US [accessed 2021-01-28]
51. Care19 Diary, ProudCrowd, LLC. Google Play. URL: https://play.google.com/store/apps/details?id=com.proudcrowd.care&hl=en_AU&gl=US [accessed 2021-01-28]
52. TraceTogether, Government Technology Agency. Google Play. URL: https://play.google.com/store/apps/details?id=sg.gov.tech.bluetrace&hl=en_AU [accessed 2021-01-28]
53. COVIDSafe app. Australian Government Department of Health. 2020. URL: https://www.health.gov.au/resources/apps-and-tools/covidsafe-app [accessed 2021-01-28]
54. Dahiya N, Kakkar AK. Mobile health: Applications in tackling the Ebola challenge. J Family Med Prim Care 2016;5(1):192-193 [FREE Full text] [doi: 10.4103/2249-4863.184667] [Medline: 27453876]

55. Rowe F, Ngwenyama O, Richet J. Contact-tracing apps and alienation in the age of COVID-19. Eur J Inf Syst 2020 Sep 13;29(5):545-562. [doi: 10.1080/0960085x.2020.1803155]
56. Morley J, Cowls J, Taddeo M, Floridi L. Ethical guidelines for COVID-19 tracing apps. Nature 2020 Jun;582(7810):29-31. [doi: 10.1038/d41586-020-01578-0] [Medline: 32467596]
57. Gu T. Newzoo. 2019 Sep 17. URL: https://newzoo.com/insights/articles/newzoos-global-mobile-market-report-insights-into-the-worlds-3-2-billion-smartphone-users-the-devices-they-use-the-mobile-games-they-play/ [accessed 2021-01-28]
58. Hutton L, Price BA, Kelly R, McCormick C, Bandara AK, Hatzakis T, et al. Assessing the Privacy of mHealth Apps for Self-Tracking: Heuristic Evaluation Approach. JMIR Mhealth Uhealth 2018;6(10) [FREE Full text] [doi: 10.2196/mhealth.9217] [Medline: 30348623]
59. Zhao Q, Wen H, Lin Z, Xuan D, Shroff N. On the accuracy of measured proximity of Bluetooth-based contact tracing apps. In: Proceedings of the International Conference on Security and Privacy in Communication Systems. 2020 Presented at: International Conference on Security and Privacy in Communication Systems; October 21-23, 2020; Washington, WA p. 49-60. [doi: 10.1007/978-3-030-63086-7_4]
60. Cho H, Ippolito D, Yu Y. Contact Tracing Mobile Apps for COVID-19: Privacy Considerations and Related Trade-offs. arXiv.org 2020:1-12 [FREE Full text]
61. Poudyal A, van Heerden A, Hagaman A, Maharjan SM, Byanjankar P, Subba P, et al. Wearable Digital Sensors to Identify Risks of Postpartum Depression and Personalize Psychological Treatment for Adolescent Mothers: Protocol for a Mixed Methods Exploratory Study in Rural Nepal. JMIR Res Protoc 2019;8(8):e14734 [FREE Full text] [doi: 10.2196/14734] [Medline: 31512581]
62. Elkhodr M, Shahrestani S, Cheung H. A semantic obfuscation technique for the Internet of Things. In: Proceedings of the International Conference on Communications Workshops (ICC). 2014 Presented at: IEEE International Conference on Communications Workshops (ICC); June 10-14, 2014; Sydney, Australia p. 448-453. [doi: 10.1109/iccw.2014.6881239]
63. Arsalan M, Mubin O, Alnajjar F, Alsinglawi B, Zaki N. Global and temporal COVID-19 risk evaluation. Front Public Health 2020;8:440 [FREE Full text] [doi: 10.3389/fpubh.2020.00440] [Medline: 32850611]

Abbreviations

API: application programming interface
DP-3T: Decentralized Privacy-Preserving Proximity Tracing
PEPP-PT: Pan-European Privacy-Preserving Proximity Tracing
QR: Quick Response
TCN: Temporary Contact Numbers

Edited by R Kukafka; submitted 13.08.20; peer-reviewed by J Li, A Johnson; comments to author 25.09.20; revised version received 14.10.20; accepted 20.01.21; published 09.02.21.

Please cite as:
Elkhodr M, Mubin O, Iftikhar Z, Masood M, Alsinglawi B, Shahid S, Alnajjar F
Technology, Privacy, and User Opinions of COVID-19 Mobile Apps for Contact Tracing: Systematic Search and Content Analysis
J Med Internet Res 2021;23(2):e23467
URL: http://www.jmir.org/2021/2/e23467/
doi:10.2196/23467
PMID:33493125

Original Paper

Early Crowdfunding Response to the COVID-19 Pandemic: Cross-sectional Study

Sameh Nagui Saleh[1,2], MD; Christoph U Lehmann[2,3], MD; Richard J Medford[1,2], MD

[1]Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX, United States

[2]Clinical Informatics Center, University of Texas Southwestern Medical Center, Dallas, TX, United States

[3]Departments of Pediatrics, Bioinformatics, Population & Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX, United States

Corresponding Author:
Sameh Nagui Saleh, MD
Department of Internal Medicine
University of Texas Southwestern Medical Center
5323 Harry Hines Blvd
Dallas, TX, 75390
United States
Phone: 1 5713383680
Email: sameh.n.saleh@gmail.com

Abstract

Background: As the number of COVID-19 cases increased precipitously in the United States, policy makers and health officials marshalled their pandemic responses. As the economic impacts multiplied, anecdotal reports noted the increased use of web-based crowdfunding to defray these costs.

Objective: We examined the web-based crowdfunding response in the early stage of the COVID-19 pandemic in the United States to understand the incidence of initiation of COVID-19–related campaigns and compare them to non–COVID-19–related campaigns.

Methods: On May 16, 2020, we extracted all available data available on US campaigns that contained narratives and were created between January 1 and May 10, 2020, on GoFundMe. We identified the subset of COVID-19–related campaigns using keywords relevant to the COVID-19 pandemic. We explored the incidence of COVID-19–related campaigns by geography, by category, and over time, and we compared the characteristics of the campaigns to those of non–COVID-19–related campaigns after March 11, when the pandemic was declared. We then used a natural language processing algorithm to cluster campaigns by narrative content using overlapping keywords.

Results: We found that there was a substantial increase in overall GoFundMe web-based crowdfunding campaigns in March, largely attributable to COVID-19–related campaigns. However, as the COVID-19 pandemic persisted and progressed, the number of campaigns per COVID-19 case declined more than tenfold across all states. The states with the earliest disease burden had the fewest campaigns per case, indicating a lack of a case-dependent response. COVID-19–related campaigns raised more money, had a longer narrative description, and were more likely to be shared on Facebook than other campaigns in the study period.

Conclusions: Web-based crowdfunding appears to be a stopgap for only a minority of campaigners. The novelty of an emergency likely impacts both campaign initiation and crowdfunding success, as it reflects the affective response of a community. Crowdfunding activity likely serves as an early signal for emerging needs and societal sentiment for communities in acute distress that could be used by governments and aid organizations to guide disaster relief and policy.

(J Med Internet Res 2021;23(2):e25429) doi:10.2196/25429

KEYWORDS

crowdfunding; fundraising; GoFundMe; COVID-19; coronavirus; pandemic; natural disasters; disaster relief; fundraise; disaster; cross-sectional; crowdfund; crowdsource; economy; social media; community; distress

Introduction

As the number of COVID-19 cases precipitously accelerated in March throughout the United States, government and public health departments marshalled their pandemic responses. Support efforts quickly became necessary as shortages of medical supplies and testing worsened by the continued spread of the virus affected health care organizations, medical offices, and nursing homes [1]. The first effects of the pandemic were related to health; however, economic ramifications, including rapidly increasing unemployment and loss of revenue, quickly followed the spread of the disease and public health efforts to contain it [2,3]. Temporary and permanent closure of small businesses, furloughing of whole sectors of the economy (eg, airline and restaurant industries), and food price inflation created social, financial, and medical strain on the population [4,5]. Subsequent effects of the pandemic, such as the increased number of deaths and associated funeral costs, added more financial stress to the lives of Americans.

Donation-based web-based crowdfunding has become an increasingly popular tool to finance medical treatment, respond to financial hardships and natural disasters, and defray the downstream economic impacts of personal health care costs [6,7]. Previous research has shown that crowdfunding has been used to address the needs of individuals and communities after natural disasters [8]. However, the crowdfunding success of these donation-based platforms relies more heavily on emotional delivery to engage backers [9] as well as on social media for quick diffusion [10].

Despite anecdotal reports of the increased use of web-based crowdfunding to mitigate the economic impacts of COVID-19 [11], details on how crowdfunding has been used for COVID-19 relief efforts in the United States remain scarce. We evaluated web-based crowdfunding campaigns in the United States in the early stages of the COVID-19 pandemic on the GoFundMe platform, which represents 90% of the social crowdfunding market in the United States as of 2018 [12]. We hypothesized that (1) COVID-19–related campaigns raised more money and were more widely distributed on social media than non–COVID-19–related campaigns during the same timeframe and (2) COVID-19–related campaigns were initiated disproportionately in the early days and weeks of the pandemic, unrelated to the rate of incident COVID-19 cases.

Methods

Data Collection and Analysis

On May 16, 2020, we extracted all active US campaigns created between January 1 and May 10, 2020, on GoFundMe, the largest web-based crowdfunding platform in the United States. The time frame reflects approximately two months before and after the World Health Organization declared COVID-19 a pandemic on March 11, 2020. We used the *Beautiful Soup* library [13] to web scrape all data available for the GoFundMe campaigns and excluded all campaigns that originated outside the United States. We list all extracted data fields in Table S1 in Multimedia Appendix 1. We used keywords relevant to the COVID-19 pandemic to identify and label the subset of COVID-19–related campaigns (Table S2 in Multimedia Appendix 1). The University of Texas Southwestern Human Research Protection Program Policies, Procedures, and Guidance did not require institutional review board approval, as all data were publicly available.

We examined weekly new COVID-19–related campaigns juxtaposed with the weekly numbers of new COVID-19 cases both nationally and by state or territory. We evaluated the incidence of campaigns by state adjusted for new COVID-19 cases by week to determine the temporal association between disease spread and campaign initiation. To enable comparison among states, we defined week 0 as the first week in which each state had at least 100 new weekly cases. We displayed all states; however, we only highlighted the top six states by overall campaign per million. Given the presence of zip code–level data, we also explored more granular community trends. We also evaluated changes in COVID-19–related campaigns per million for each state and campaign categories over time. To better understand campaign narratives and motivation, we conducted a natural language processing (NLP) analysis of the campaign text to identify clusters of keywords that represent campaigns [14]. We used the term frequency–inverse document frequency (TF-IDF) technique to create a matrix of all of the words in the campaign narratives weighted by the frequency of the word in that document relative to the frequency of the same word in all documents [15]. We then used k-means clustering on the resulting matrix to identify clusters of campaigns and represent these clusters by the 20 most central words to that cluster [16]. Because there were 19 GoFundMe categories in the COVID-19–related campaign subset (out of 20 possible GoFundMe categories), we chose 19 topic clusters a priori.

We compared campaign characteristics, including fundraising, campaign information, and campaign category, for COVID-19–related and non–COVID-19–related campaigns over the same time period from March 11 to May 10, 2020. We used Mann-Whitney U, chi-square, and Fisher exact tests where appropriate to determine significance. The α level of significance was set a priori at .05, and all significance testing was two-sided. We did not adjust for multiple comparisons, as this was an exploratory study and should be interpreted as hypothesis-generating. Analyses were performed using Python, version 3.7.2 (Python Software Foundation).

Availability of Data and Materials

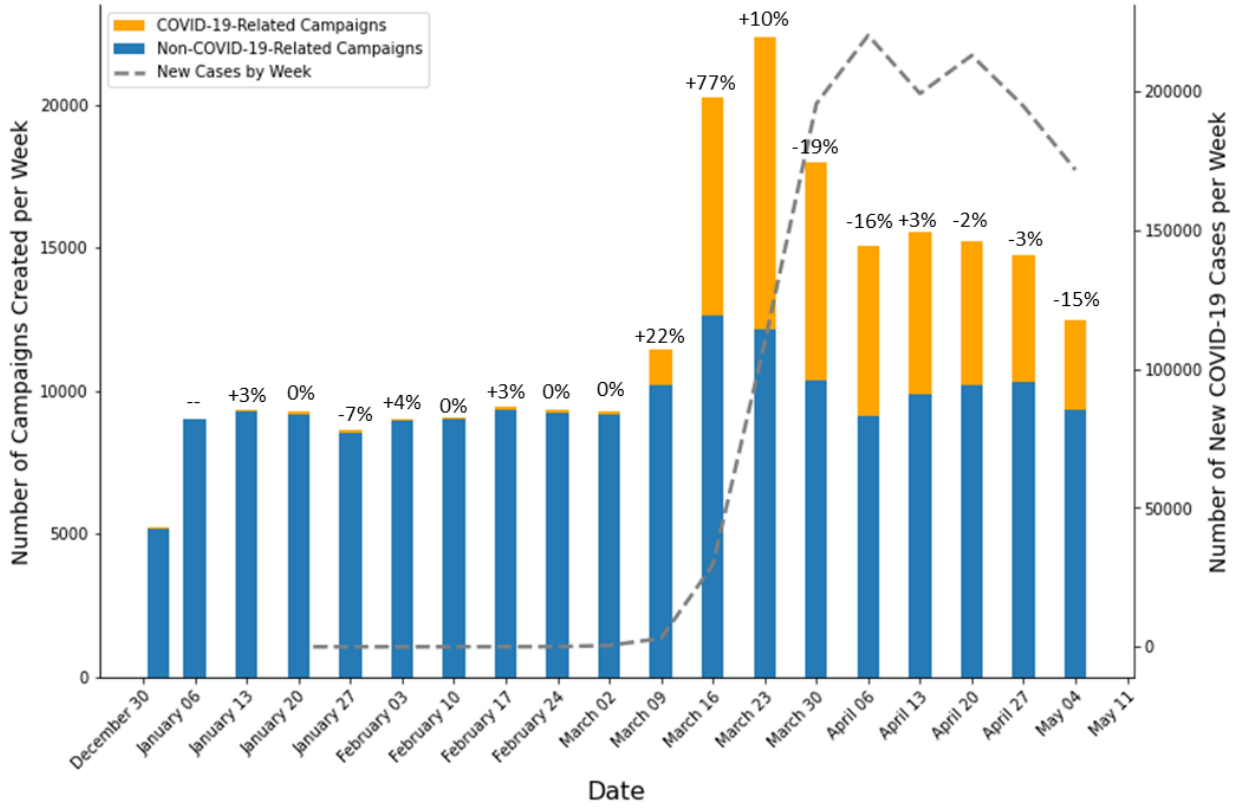
The data that support the findings of this study are available upon request.

Results

Of 310,695 total new campaigns created between January 1 and May 10, 2020, we analyzed 232,827 campaigns from the United States across all categories. Of these campaigns, 51,763/232,827 (22.2%) were identified as COVID-19–related, and these campaigns collectively raised US $237,418,235 by May 10, 2020. The vast majority of COVID-19–related campaigns were established after March 11 (50,828/51,763, 98.2%). New GoFundMe campaigns peaked in mid-March, mainly due to new campaigns related to or referencing COVID-19, and

declined in April (Figure 1). Total campaigns increased by 77% and 10% in the weeks of March 16 and 23 and then dipped by 19% and 16% in the following two weeks. In parallel, COVID-19–related campaigns increased by 736% and 512% in the weeks of March 9 and 16 and then eventually decreased by 25% and 23% in the weeks of March 30 and April 6. The COVID-19–related increase in campaigns was observed before the national peak of new COVID-19 cases in April. Subsequently, the incidence of new campaigns declined, while the incidence of new COVID-19 cases remained steady or declined slightly.

Figure 1. New GoFundMe campaigns by week versus incident COVID-19 cases by week. Percent changes of campaigns created from the previous week are shown above each bar.



Adjusting for the weekly incidence of COVID-19 cases, new COVID-19–related campaigns in all US states peaked within two weeks of the first week when 100 new cases were diagnosed in that state. For many states, the rate of COVID-19–related campaigns per case dropped by at least tenfold after the first two to three weeks (Figure 2). Population-rich states that also experienced early spread of COVID-19, such as New York, New Jersey, Massachusetts, Washington, and Illinois, had the most COVID-19–related campaigns per million inhabitants (Figure 3); however, they were ultimately among the states with the lowest number of COVID-19–related campaigns per 1000 cases during the study period.

Figure 2. COVID-19–related GoFundMe campaigns per 1000 COVID-19 incident cases for each state. Week 0 is defined as the first week in which a state had more than 100 cases to enable direct comparison among states. All states are shown, but only the top 6 states for total COVID-19 cases per million are highlighted in color (ordered from most to fewest in the legend).

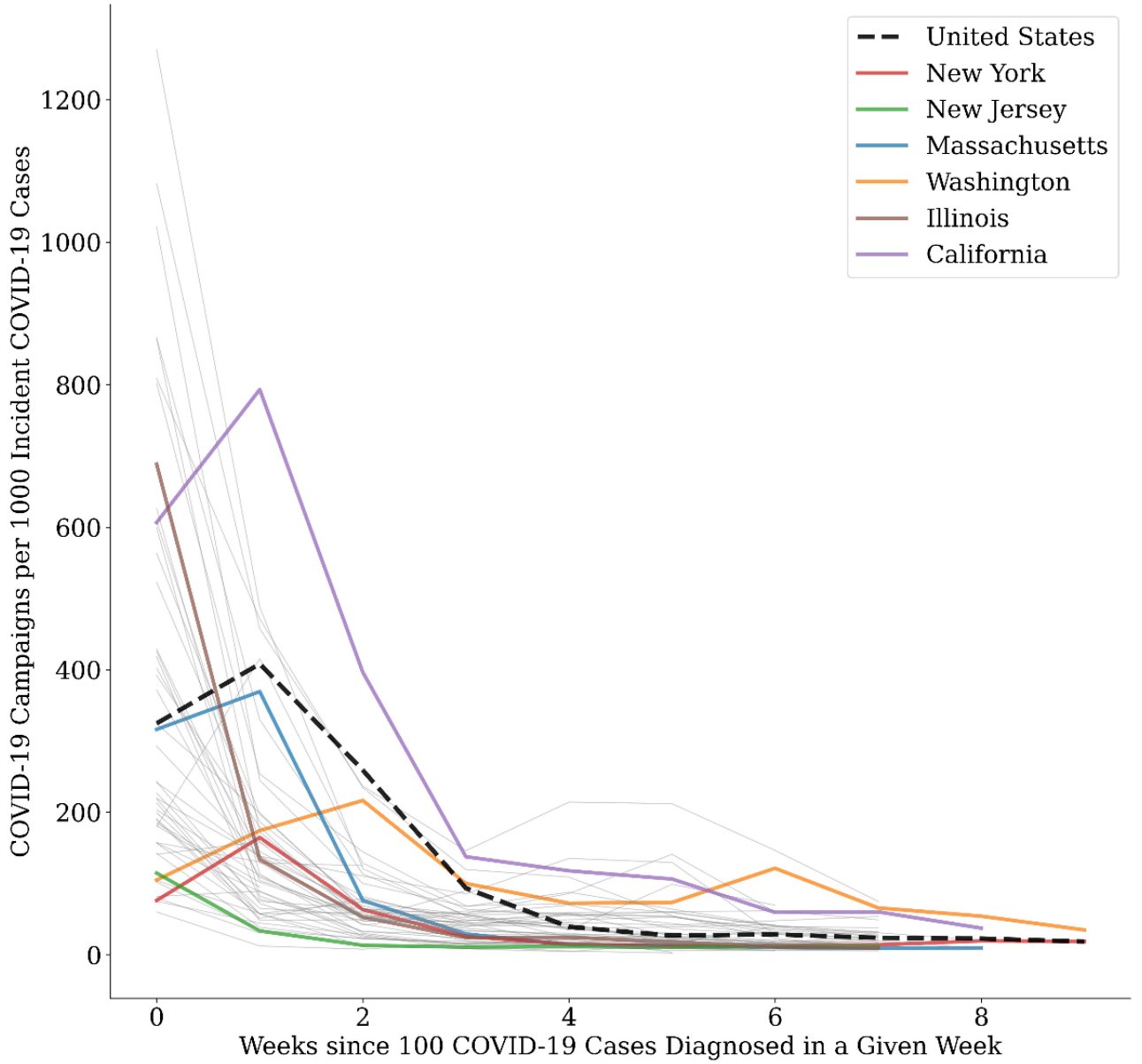
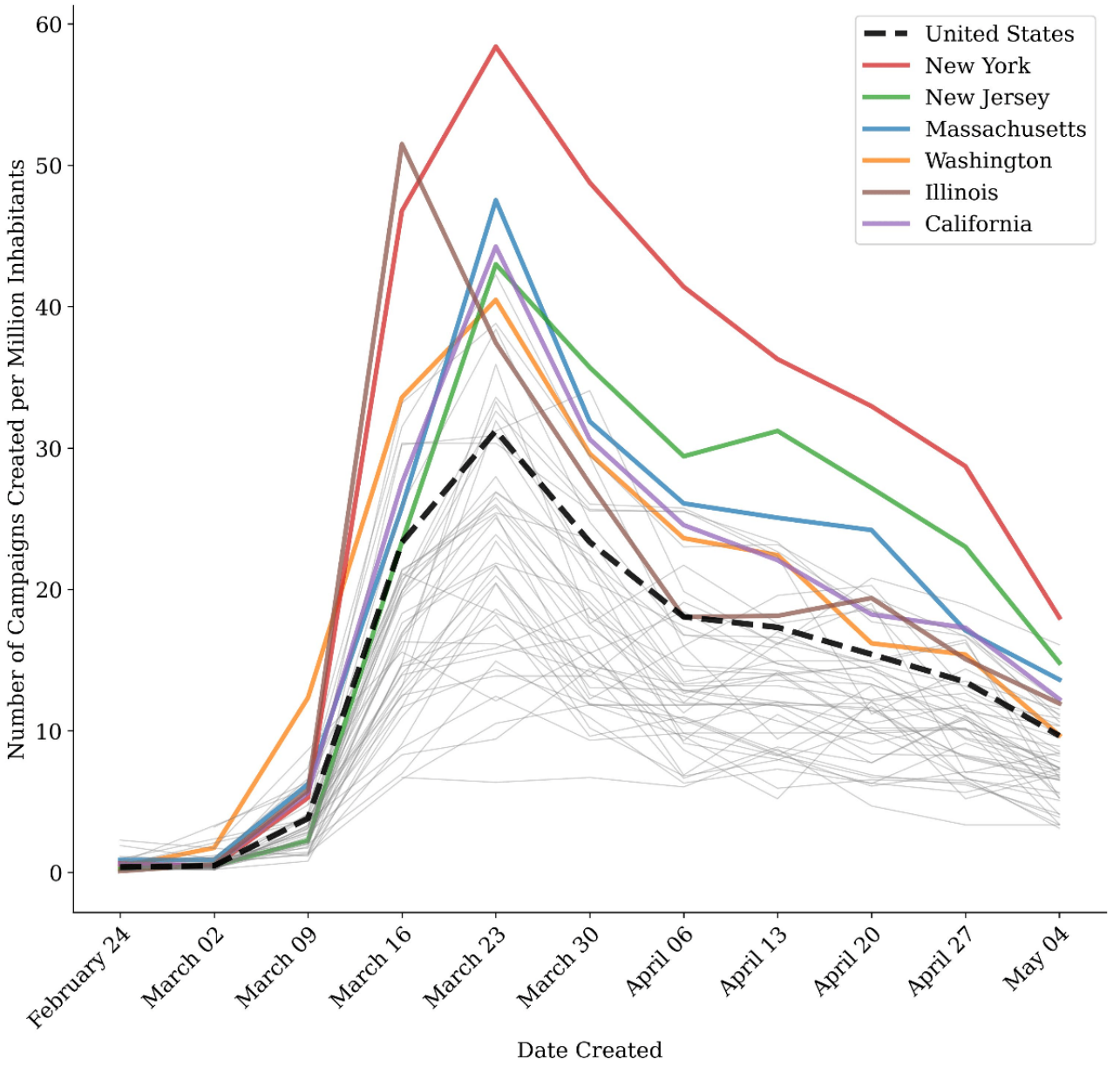


Figure 3. COVID-19–related GoFundMe campaigns per million by week by state. All states are shown, but only the top 6 states for total COVID-19 cases per million are highlighted in color (ordered from most to fewest in the legend). Dates on the x-axis start from February 25, as few COVID-19–related campaigns were present before this date.



From March 11 to May 10, 2020, non-COVID-19–related campaigns (91,631/142,459, 64.3%) raised a median of US $625 (IQR $135-$2300), while COVID-19–related campaigns (50,828/142,459, 35.7%) raised a median of $930 (IQR $220-$3075), nearly 50% more per campaign (*P*<.001). Fundraising goals were higher in COVID-19–related campaigns (median $5000, IQR $2000-$10,000) than non–COVID-19–related campaigns (median $4000, IQR $1250-$10,000; *P*<.001). Even with the higher fundraising goals, COVID-19–related campaigns (median 25.0%, IQR 5.5%-69.5%) raised a higher percentage of their fundraising goals than non–COVID-19–related campaigns (median 22.2%, IQR 5.0%-65.5%) by date of data extraction (*P*<.001). COVID-19–related campaigns (median 23, IQR 0-163) had significantly more Facebook shares than non–COVID-19–related campaigns (median 0, IQR 0-145; *P*<.001). COVID-19–related campaigns also had longer narratives in their campaign description, were more likely to be listed as a charity, and were more likely to have the campaigner be the same as the beneficiary (Table 1).

Table 1. Baseline GoFundMe campaign characteristics stratified by COVID-19–related status. Only campaigns created after March 11, 2020, are included.

Characteristic	Total campaigns (N=142,459)	COVID-19–related campaigns (n=50,828)	Non–COVID-19–related campaigns (n=91,631)	*P* value[a]
Fundraising				
Goal (US $), median (IQR)	5000 (1500-10,000)	5000 (2000-10,000)	4000 (1250-10,000)	<.001
Raised (US $), median (IQR)	722 (160-2,565)	930 (220-3075)	625 (135-2300)	<.001
Percent of goal funded (US $), median (IQR)	23.3 (5.2-66.9)	25.0 (5.5-69.5)	22.2 (5.0-65.5)	<.001
Donors, median (IQR)	12 (3-35)	14 (4-40)	11 (3-32)	<.001
Amount per donation (US $), median (IQR)	56.3 (35.4-86.8)	59.9 (39.0-92.7)	54.5 (33.8-83.5)	<.001
Met funding goal, n (%)	20,593 (14.5)	7302 (14.4)	13,291 (14.5)	.48
Campaign information				
Campaigner = beneficiary, n (%)	119,243 (83.7)	43,445 (85.5)	75,798 (82.7)	<.001
Is listed as a charity, n (%)	7144 (5.0)	3629 (7.1)	3515 (3.8)	<.001
Words in narrative, median (IQR)	145 (84-240)	187 (115-299)	125 (72-207)	<.001
Characters in narrative, median (IQR)	56.3 (35.4-86.8)	1072 (652-1729)	691 (393-1154)	<.001
Active donation days, median (IQR)	23 (14-33)	24 (15-38)	22 (13-31)	<.001
Facebook shares, median (IQR)	4 (0-152)	23 (0-163)	0 (0-145)	<.001
GoFundMe hearts, median (IQR)	12 (3-33)	14 (4-38)	10 (3-31)	<.001
Campaign category (top 8 of 20)				
Accidents & Emergencies, n (%)	25,306 (17.8)	10,705 (21.1)	14,601 (15.9)	<.001
Medical, Illness & Healing, n (%)	22,669 (15.9)	8484 (16.7)	14,185 (15.5)	<.001
Funerals & Memorials, n (%)	19,577 (13.7)	3242 (6.4)	16,335 (17.8)	<.001
Community & Neighbors, n (%)	14,237 (10.0)	7008 (13.8)	9027 (7.9)	<.001
Business & Entrepreneurs, n (%)	11,451 (8.0)	6458 (12.7)	7229 (5.4)	<.001
Animals & Pets, n (%)	11,158 (7.8)	2131 (4.2)	9027 (9.9)	<.001
Babies, Kids, & Family, n (%)	6265 (4.4)	1990 (3.9)	4275 (4.7)	<.001
Volunteer & Service, n (%)	5176 (3.6)	2643 (5.2)	2533 (2.8)	<.001

[a]Given the non-Gaussian populations, we used Mann-Whitney U, chi-square, and Fisher exact tests to detect statistical differences among the three groups.

Creators of COVID-19–related campaigns most commonly selected the following categories (in descending order by number of campaigns): (1) Accidents & Emergencies, (2) Medical, Illness & Healing, (3) Community & Neighbors, (4) Business & Entrepreneurs, (5) Funerals & Memorials. In total, 19 of the 20 possible GoFundMe categories were represented in the COVID-19–related campaigns. Accidents & Emergencies as well as Community & Neighbors and Medical, Illness & Healing had the earliest peaks (Figure 4A). A week later, the category Business & Entrepreneurs peaked; however, this appeared to be driven by an isolated spike on March 24 (Figure 4B). Medical, Illness & Healing peaked next within the following 2 weeks. As these initial spikes decreased significantly, Funerals & Memorials campaigns lagged behind the first peak by 4 weeks. Compared to non-COVID-19–related campaigns after March 11, COVID-19–related campaigns had more campaigns in Accidents & Emergencies, Community & Neighbors, and Business and Entrepreneurs and fewer campaigns in Funerals & Memorials and Animals & Pets. Our NLP analysis provides granular insight into motivators for campaigns (Table 2). The largest cluster (Cluster 0) was about community support and fundraising for schools and children in a time of need. There are also multiple clusters about health care workers and personal protective equipment (Clusters 7, 11, and 18), small businesses and jobs (Clusters 3, 5, 6, and 9), as well as funerals and medical costs (Clusters 8, 15, and 17). We also observed the clustering of likely Spanish-only campaigns in Cluster 12. Cluster 19, with only 3 campaigns, likely accounts for outliers among the other campaigns and is included for completeness.

Figure 4. COVID-19–related GoFundMe campaigns per category (A) by week and (B) by day. Only the top 8 of the 20 possible categories are shown. The legend is ordered from most to least common category. Dates on the x-axis start from February 25, as few COVID-19–related campaigns were present before this date.

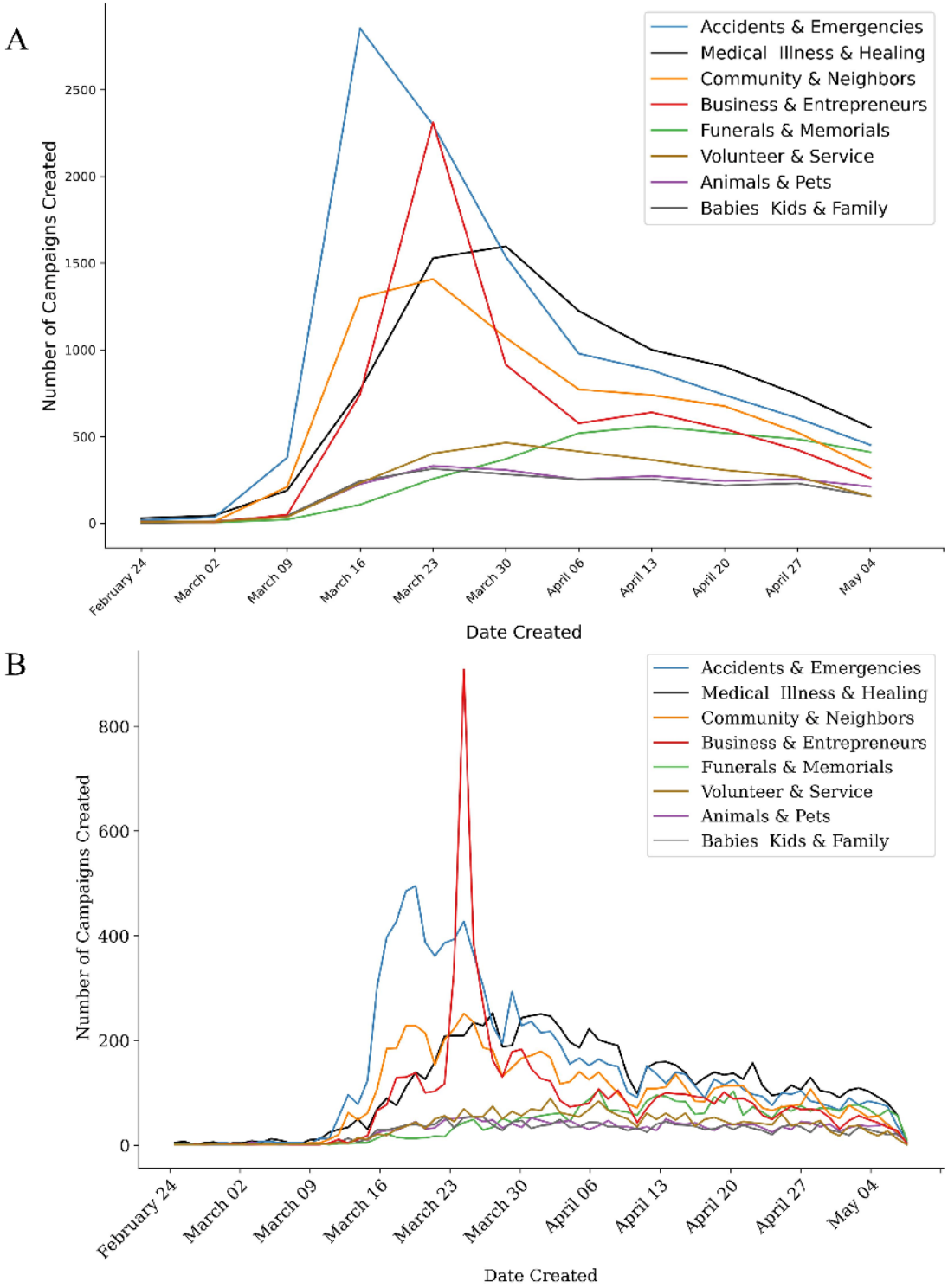


Table 2. Keywords at the center of the 19 clusters of campaigns obtained through the term frequency–inverse document frequency natural language processing technique and k-means clustering. The top 20 words for each cluster and the number of campaigns per cluster are displayed.

Cluster	Number of campaigns	Top 20 words in the cluster
1	10,421	community, school, support, students, time, need, pandemic, children, money, funds, thank, donation, make, continue, people, like, year, new, donations, teachers
2	6493	know, time, family, home, work, just, thank, life, need, able, baby, love, cancer, going, like, money, years, little, pandemic, people
3	4032	business, small, businesses, salon, support, open, time, community, smallbusinessrelief, clients, doors, rent, thank, able, stay, family, years, close, like, continue
4	3785	families, people, food, need, money, community, family, support, provide, supplies, time, pandemic, children, funds, feed, affected, work, basic, make, world
5	3715	job, work, pay, rent, money, just, corona, virus, car, bills, need, know, time, thank, able, family, really, lost, going, well
6	3171	staff, support, bar, time, doors, community, restaurant, family, thank, love, bartenders, know, team, members, employees, work, close, industry, times, soon
7	2935	healthcare, workers, medical, hospitals, patients, hospital, ppe, nurses, care, supplies, health, masks, equipment, protective, need, doctors, support, local, lives, working
8	2776	family, father, funeral, passed, away, life, loved, time, dad, 2020, brother, april, expenses, wife, husband, thank, know, lost, love, friend
9	2613	employees, team, restaurant, support, time, community, staff, family, doors, business, thank, customers, members, industry, restaurants, fund, work, close, crisis, open
10	2530	food, meals, local, community, restaurants, workers, support, meal, feed, provide, need, people, healthcare, families, hospital, donations, feeding, pandemic, money, responders
11	1567	masks, mask, hospitals, n95, workers, healthcare, medical, need, face, supplies, ppe, fabric, make, materials, surgical, sewing, protective, donate, hospital, purchase
12	1553	que, en, la, el, para, por, su, los, se, una, es, lo, mi, las, esta, del, familia, sus, ayuda, todos
13	1320	artists, industry, workers, fund, musicians, relief, support, community, music, service, work, people, need, funds, income, time, restaurant, financial, pandemic, assistance
14	1187	surgery, vet, dog, pain, just, life, time, know, needs, leg, family, right, old, emergency, hospital, work, going, need, cancer, tumor
15	1175	hospital, family, medical, bills, ventilator, time, blood, home, days, care, icu, prayers, thank, know, heart, life, doctors, weeks, positive, april
16	871	difference, contribution, benefit, cause, join, impact, raising, advance, thanks, making, means, donation, make, money, want, information, food, america, foundation, community
17	867	mother, family, mom, time, lost, funeral, life, thank, home, sister, daughter, know, children, son, passed, single, years, friend, work, care
18	749	shields, face, 3d, shield, ppe, printing, printer, printers, materials, masks, workers, medical, hospitals, healthcare, filament, produce, print, equipment, production, need
19	3	quarnatines, exposure, 14, response, self, support, consider, donating, days, stay, want, virus, home, family, time, alejandro, eua, eubank, eubanks, eu

Discussion

Principal Findings

Early in the COVID-19 pandemic, COVID-19–related campaigns were created more frequently. There was a substantial increase in overall GoFundMe web-based crowdfunding campaigns in March, mostly attributable to COVID-19–related campaigns. COVID-19–related campaigns raised more money than other campaigns, had longer narrative descriptions, and were more likely to be shared on Facebook. However, as the COVID-19 pandemic progressed, we suspect that the novelty of these campaigns wore off. As COVID-19 became a familiar and ubiquitous problem, the number of campaigns per COVID-19 case declined across all states. As proposed by Elmer et al [17], we agree that crowdfunding serves as a lens for the affective responses to the pandemic, showing both the economic uncertainties and the anxieties within a community.

We speculate that as in commercial advertising, the creativity and novelty of a campaign as well as its quick diffusion play major roles in the campaign's success, enabling it to progress from "friend and family funding" to crowdfunding. The subsequently declining number of COVID-19–related campaigns may be explained by potential campaign developers weighing the effort required to create a campaign with a shrinking potential for reward. Initially, the novelty and early shock of COVID-19 cases provided the chance for a campaign to "go

viral"; reach many people, especially through social media; and generate large donations. Social media and mobilizing users outside the immediate family and friend network are key to attract the attention of a larger group of funders [18] and achieve crowdfunding success [19]. In fact, the use of social media has been shown to predict the volume of emergency relief donations in cases of natural disasters more accurately than conventional techniques, highlighting the importance of quick diffusion [10]. As the issue became widespread and the ability to set a campaign apart as a worthy cause became more difficult, the effort (cost) of creating a campaign may have been outweighed by the perceived likelihood of funding. As others have found, the concept of worthiness and augmenting one's "illness narrative" become important factors in generating campaign appeal, influence, and ultimately fundraising success [6,20-22]. The states with the earliest disease burden had the fewest campaigns per case, indicating a lack of case-dependent response; this supports the premise that potential campaign designers perceive common illnesses as less likely to receive the attention of donors over time.

We recognize that the impetus for crowdfunding is a complex interplay of many factors. Campaigns may be generated for an individual, multiple individuals, an institution such as a business, or a larger community. External factors such as government readiness, supply shortages, and lockdowns may play roles in campaign incidence. However, the rapid rise and fall in campaigns over a timeline of two or three weeks in nearly all states points to a reactionary, affective community response; it also indicates that the external factors were not meaningfully reversed or corrected during this time. The gravity and magnitude of the event likely sparked reactions of desperation, fear, and anticipation that motivated campaigners to turn reaction to action, especially with the possibility of the absence of other immediate alternatives early on.

Although web-based crowdfunding raises millions of dollars, our study suggests that it may function as a financial safety net for a limited subset of beneficiaries. First, there is ample evidence that crowdfunding success disproportionately benefits those in areas with high socioeconomic status and those with the internet and media literacy necessary to portray beneficiaries as worthy [6,21-23]. Second, medical expenses alone due to COVID-19 in the United States are estimated to be in excess of $163 billion if 20% of the population is infected (on August 26, 2020, approximately 1.8% of the population had a confirmed infection) [24]. Rather, crowdfunding more likely functions as a weathervane indicating a community in distress. As Elmer et al [17] noted, "while clearly producing an economic crisis for governments, communities, families, companies and individuals, the [COVID-19] crisis is rooted in issues of community health." Further, crowdfunding is heavily swayed by marketing and storytelling, as it points a majority of funds toward a few cases (when these elements converge to virality) rather than aiding the many [11,25].

The large number of categories for COVID-19–related GoFundMe campaigns tells a poignant story of the broad, destructive effect that COVID-19 has had on society. Beyond creating the need to cover unexpected and costly medical expenses, the COVID-19 pandemic devastated small businesses, greatly increased unemployment, created food and housing insecurities, generated unexpected expenses such as funeral costs, and increased debt [5]. The campaign categories reflected the progression of the COVID-19 funding response, with the earliest peaks for accidents and emergencies and late growth and plateaus for funerals and memorials. Our NLP analysis provides further evidence of the content of the campaigns, as the clusters identified align with the specified categories listed.

The spike in small business campaigns was likely heavily based around the launch of the Small Business Relief Initiative on March 24, 2020 [26], and specifically, its GoFundMe.org (the charitable arm of GoFundMe) Small Business Relief Fund. The relief fund provided "one-time $500 matching grants to qualifying small businesses that created a fundraiser through the Small Business Relief Initiative or already had an existing GoFundMe until the relief fund [was] depleted." For "fundraisers started prior to this announcement or outside of this partnership, the organizer [could] update their fundraiser description with the hashtag #SmallBusinessRelief to receive funds," which likely accounted for the presence of this phrase in Cluster 3 in our NLP analysis. Importantly, another part of the initiative involved automatically creating fundraisers for small businesses with Yelp pages, where customers could donate directly on Yelp or on GoFundMe; however, because we only scraped GoFundMe-based campaigns that included a user-generated narrative, these fundraisers were not part of our dataset.

As newly generated GoFundMe campaigns showed a temporal association with the effect of the COVID-19 pandemic on society, we contemplated that GoFundMe campaigns may serve as an early indicator of a community in distress. To explore this, we isolated campaigns from communities that were affected by a natural disaster, including Puerto Rico (earthquake from January 6-7, 2020), Nashville (tornado from March 2-3, 2020), and Mississippi (tornado from April 12-13, 2020). We were able to discern an immediate peak for GoFundMe campaigns following each event (Figure S1, Multimedia Appendix 1).

Although GoFundMe campaigns failed to meet the needs of those seeking support, as demonstrated by the small fraction of funding goals met by the data extraction date, the signal of new campaigns at and after times of distress can provide unique insight into the community affected. The organic, user-generated content of these campaigns can inform both direct emerging needs (whether resource-related or financial) as well as the societal perception and sentiment in a time of crisis [17]. This signal does not appear to be limited to medical emergencies but also extends to natural disasters. Future research will show which types of negative events that affect communities are reflected by a spike in GoFundMe campaigns and how the content of the campaigns can be leveraged systematically.

Limitations

Although we were able to analyze the full available set of GoFundMe campaigns, which strengthened our study, it had several limitations. Web-based crowdfunding only provides a lens to view those who have access to the technical and social resources to be a part of a crowdfunding campaign (either to

create and coordinate a campaign directly or to be connected to another person who can fulfill that role). Therefore, only a subset of the population experiencing unmet financial strain or needs is represented. Moreover, we used data from one crowdfunding platform only. Although GoFundMe is the largest platform in the United States by number of campaigns, our study may have suffered from selection bias. Because of the large number of campaigns, we are unable to validate that campaigns using COVID-19–related keywords were indeed designed to mitigate the effect of COVID-19. Although we note significant differences in campaign and funding characteristics between COVID-19–related and non–COVID-19–related campaigns, future work will be important to further understand reasons for these differences and their implications in long-term crowdfunding. Finally, we only analyzed the first few months of COVID-19–related campaigns; therefore, more longitudinal analysis will be important to understand more completely how peaks and troughs in COVID-19 incidence nationally and in more granular communities affect campaign frequency.

Conclusions

Web-based crowdfunding activity leverages galvanized public reaction early in a public health emergency such as the COVID-19 pandemic. However, as disease spread persisted and economic burden continued to grow nationally, the creation of new campaigns faded. Web-based crowdfunding appears to be a stopgap for only a minority of campaigners. The novelty of an emergency likely impacts both campaign initiation and crowdfunding success, as it reflects the affective response of a community.

More importantly than its rather weak effect of mitigating the effect of the crisis, crowdfunding activity likely provides an early signal of the emerging needs and societal sentiment of communities in acute distress. This activity could be monitored by local, state, and federal government agencies, such as the Federal Emergency Management Association (FEMA), to provided targeted support and policy. Future research will determine what emergencies generate the strongest signals in crowdfunding activities and how content can be leveraged systematically.

Authors' Contributions

SS and RJM conceived and designed the study and acquired the data. SNS analyzed and interpretated the data. SNS and CUL prepared the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

CUL reports stock ownership in Celanese Corporation and Colfax Corporation. RJM has received research funding from Verily Life Sciences and research grants from the Centers for Disease Control and Prevention. The remaining author has no conflicts to declare.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 526 KB - jmir_v23i2e25429_app1.docx]

References

1. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair allocation of scarce medical resources in the time of Covid-19. N Engl J Med 2020 May 21;382(21):2049-2055. [doi: 10.1056/nejmsb2005114]
2. Almost 25 million jobs could be lost worldwide as a result of COVID-19, says ILO. International Labour Organization. 2020 Mar 18. URL: https://www.ilo.org/global/about-the-ilo/newsroom/news/WCMS_738742/lang--en/index.htm [accessed 2020-08-24]
3. Tappe A. 30 million Americans have filed initial unemployment claims since mid-March. CNN Business. 2020 Apr 30. URL: https://www.cnn.com/2020/04/30/economy/unemployment-benefits-coronavirus/index.html [accessed 2020-07-01]
4. Sherwan, Erik. Coronavirus wreaks havoc on Chinese food producers, sending inflation soaring. Fortune. 2020 Feb 15. URL: https://fortune.com/2020/02/15/china-coronavirus-business-impact-inflation-economy/ [accessed 2020-07-13]
5. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): a review. Int J Surg 2020 Jun;78:185-193 [FREE Full text] [doi: 10.1016/j.ijsu.2020.04.018] [Medline: 32305533]
6. Saleh SN, Ajufo E, Lehmann CU, Medford RJ. A comparison of online medical crowdfunding in Canada, the UK, and the US. JAMA Netw Open 2020 Oct 01;3(10):e2021684 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.21684] [Medline: 33104206]
7. Bassani G, Marinelli N, Vismara S. Crowdfunding in healthcare. J Technol Transf 2018 Apr 20;44(4):1290-1310. [doi: 10.1007/s10961-018-9663-7]
8. Radu M, McManus L. Bridging social capital through the techno-subsystem: a qualitative analysis of GoFundMe requests for hurricane relief. J Family Strengths 2019;19(1):Article 9 [FREE Full text] [doi: 10.1017/cbo9780511815447.003]

9. Rhue L, Robert L. Emotional Delivery in Pro-social Crowdfunding Success. In: CHI EA '18: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems.: ACM; 2018 Presented at: 2018 CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montréal, QC p. 1-6. [doi: 10.1145/3170427.3188534]
10. Korolov R, Peabody J, Lavoie A, Das S, Magdon-Ismail M, Wallace W. Predicting charitable donations using social media. Soc Netw Anal Min 2016 Jun 6;6(1):31. [doi: 10.1007/s13278-016-0341-1]
11. Nathaniel P, Taylor L. GoFundMe Confronts Coronavirus Demand. The New York Times. 2020 Mar 26. URL: https://www.nytimes.com/2020/03/26/style/gofundme-coronavirus.html [accessed 2020-07-13]
12. Ainsley H. GoFundMe keeps gobbling up competitors, says it's "very good for the market". Fast Company. 2018 Apr 04. URL: https://www.fastcompany.com/40554199/gofundme-keeps-gobbling-up-competitors-says-its-very-good-for-the-market [accessed 2021-02-03]
13. Beautiful Soup. Crummy: The Site. URL: https://www.crummy.com/software/BeautifulSoup/bs4/ [accessed 2021-02-03]
14. Lane H, Howard C, Hapke H. Natural Language Processing in Action: Understanding, Analyzing, and Generating Text with Python. Shelter Island, NY: Manning Publications Co; 2019.
15. Sammut C, Webb GI, editors. TF–IDF. In: Encyclopedia of Machine Learning. Boston, MA: Springer; 2011.
16. Likas A, Vlassis N, J. Verbeek J. The global k-means clustering algorithm. Pattern Recognit 2003 Feb;36(2):451-461. [doi: 10.1016/s0031-3203(02)00060-2]
17. Elmer G, Ward-Kimola S, Burton AG. Crowdfunding during COVID-19: An international comparison of online fundraising. FM 2020 Oct 16. [doi: 10.5210/fm.v25i11.10869]
18. Borst I, Moser C, Ferguson J. From friendfunding to crowdfunding: relevance of relationships, social media, and platform activities to crowdfunding performance. New Media Soc 2018 Apr;20(4):1396-1414 [FREE Full text] [doi: 10.1177/1461444817694599] [Medline: 30581357]
19. Di PF, Spagnoletti P, Prencipe A. Fundraising across digital dividevidences from charity crowdfunding. In: Lazazzara A, Nacamulli R, Rossignoli C, Za S, editors. Organizing for Digital Innovation. Cham, Switzerland: Springer; 2019:111-124.
20. Suzanne W. American health care tragedies are taking over crowdfunding. Bloomberg. 2017 Jun 12. URL: https://www.bloomberg.com/news/articles/2017-06-12/america-s-health-care-crisis-is-a-gold-mine-for-crowdfunding [accessed 2021-02-03]
21. Berliner LS, Kenworthy NJ. Producing a worthy illness: Personal crowdfunding amidst financial crisis. Soc Sci Med 2017 Aug;187:233-242. [doi: 10.1016/j.socscimed.2017.02.008] [Medline: 28274601]
22. Lukk M, Schneiderhan E, Soares J. Worthy? Crowdfunding the Canadian health care and education sectors. Can Rev Sociol 2018 Aug 09;55(3):404-424. [doi: 10.1111/cars.12210] [Medline: 29984886]
23. Silver ER, Truong HQ, Ostvar S, Hur C, Tatonetti NP. Association of Neighborhood Deprivation Index With Success in Cancer Care Crowdfunding. JAMA Netw Open 2020 Dec 01;3(12):e2026946. [doi: 10.1001/jamanetworkopen.2020.26946] [Medline: 33270122]
24. Bartsch SM, Ferguson MC, McKinnell JA, O'Shea KJ, Wedlock PT, Siegmund SS, et al. The potential health care costs and resource use associated with COVID-19 in the United States. Health Aff 2020 Jun 01;39(6):927-935. [doi: 10.1377/hlthaff.2020.00426]
25. Sumin L, Vili L. New digital safety net or just more ‘friendfunding’? Institutional analysis of medical crowdfunding in the United States. OSF. Preprint posted online on April 6, 2020. [doi: 10.31219/osf.io/9kecq]
26. GoFundMe Launches Relief Initiative to Help Small Business Impacted By COVID-19 in Partnership with Yelp. Business Wire. 2020 Mar 24. URL: https://www.businesswire.com/news/home/20200324005364/en/GoFundMe-Launches-Relief-Initiative-Small-Business-Impacted [accessed 2020-07-13]

Abbreviations

FEMA: Federal Emergency Management Association
NLP: natural language processing
TF-IDF: term frequency–inverse document frequency

Edited by G Eysenbach; submitted 01.11.20; peer-reviewed by M Lukk, J Olsen, I Idris; comments to author 23.11.20; revised version received 23.12.20; accepted 29.01.21; published 09.02.21.

Please cite as:
Saleh SN, Lehmann CU, Medford RJ
Early Crowdfunding Response to the COVID-19 Pandemic: Cross-sectional Study
J Med Internet Res 2021;23(2):e25429
URL: http://www.jmir.org/2021/2/e25429/
doi:10.2196/25429
PMID:33523826

Original Paper

Dynamic Panel Data Modeling and Surveillance of COVID-19 in Metropolitan Areas in the United States: Longitudinal Trend Analysis

Theresa B Oehmke[1], MSc; Lori A Post[2], PhD; Charles B Moss[3], PhD; Tariq Z Issa[4], BA; Michael J Boctor[4], BSc; Sarah B Welch[2], MPH; James F Oehmke[2], PhD

[1]Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, CA, United States

[2]Department of Emergency Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[3]Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, United States

[4]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Corresponding Author:
Theresa B Oehmke, MSc
Department of Civil and Environmental Engineering
University of California, Berkeley
202 O'Brien Hall
Berkeley, CA, 94720
United States
Phone: 1 5108986406
Email: toehmke@berkeley.edu

Abstract

Background: The COVID-19 pandemic has had profound and differential impacts on metropolitan areas across the United States and around the world. Within the United States, metropolitan areas that were hit earliest with the pandemic and reacted with scientifically based health policy were able to contain the virus by late spring. For other areas that kept businesses open, the first wave in the United States hit in mid-summer. As the weather turns colder, universities resume classes, and people tire of lockdowns, a second wave is ascending in both metropolitan and rural areas. It becomes more obvious that additional SARS-CoV-2 surveillance is needed at the local level to track recent shifts in the pandemic, rates of increase, and persistence.

Objective: The goal of this study is to provide advanced surveillance metrics for COVID-19 transmission that account for speed, acceleration, jerk and persistence, and weekly shifts, to better understand and manage risk in metropolitan areas. Existing surveillance measures coupled with our dynamic metrics of transmission will inform health policy to control the COVID-19 pandemic until, and after, an effective vaccine is developed. Here, we provide values for novel indicators to measure COVID-19 transmission at the metropolitan area level.

Methods: Using a longitudinal trend analysis study design, we extracted 260 days of COVID-19 data from public health registries. We used an empirical difference equation to measure the daily number of cases in the 25 largest US metropolitan areas as a function of the prior number of cases and weekly shift variables based on a dynamic panel data model that was estimated using the generalized method of moments approach by implementing the Arellano-Bond estimator in R.

Results: Minneapolis and Chicago have the greatest average number of daily new positive results per standardized 100,000 population (which we refer to as speed). Extreme behavior in Minneapolis showed an increase in speed from 17 to 30 (67%) in 1 week. The jerk and acceleration calculated for these areas also showed extreme behavior. The dynamic panel data model shows that Minneapolis, Chicago, and Detroit have the largest persistence effects, meaning that new cases pertaining to a specific week are statistically attributable to new cases from the prior week.

Conclusions: Three of the metropolitan areas with historically early and harsh winters have the highest persistence effects out of the top 25 most populous metropolitan areas in the United States at the beginning of their cold weather season. With these persistence effects, and with indoor activities becoming more popular as the weather gets colder, stringent COVID-19 regulations will be more important than ever to flatten the second wave of the pandemic. As colder weather grips more of the nation, southern metropolitan areas may also see large spikes in the number of cases.

(J Med Internet Res 2021;23(2):e26081) doi:10.2196/26081

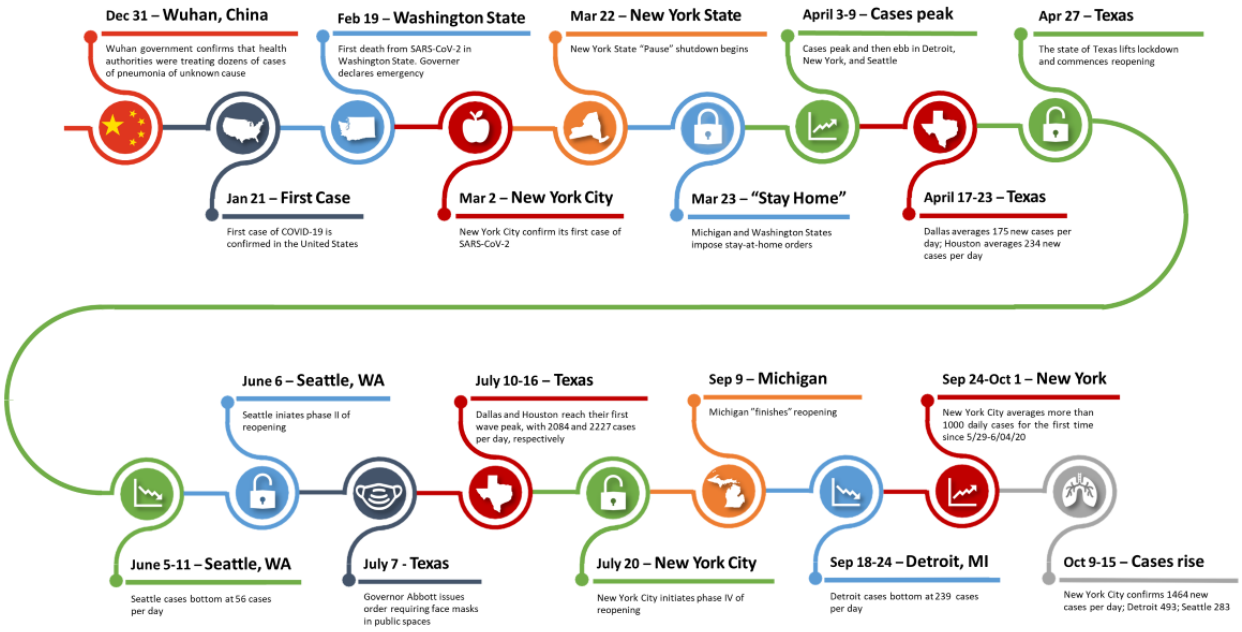
KEYWORDS

COVID-19; SARS-CoV-2; SARS-CoV-2 surveillance; second wave; wave two; wave 2; global COVID-19 surveillance; COVID-19 metropolitan areas; COVID-19 cities; US public health surveillance; US COVID-19; US surveillance metrics; dynamic panel data; generalized method of the moments; US econometrics; US SARS-CoV-2; US COVID-19 surveillance system; US COVID-19 transmission speed; US COVID-19 transmission acceleration; COVID-19 transmission deceleration; COVID-19 transmission jerk; COVID-19 7-day lag; Arellano-Bond estimator; generalized method of moments; GMM; New York City; Los Angeles; Chicago; Dallas; Houston; Washington, DC; Miami; Philadelphia; Atlanta; Phoenix; Boston; San Francisco; Riverside; Detroit; Seattle; Minneapolis; San Diego; Tampa; Denver; St Louis; Baltimore; Charlotte; Orlando; San Antonio; Portland

Introduction

In December 2019, a novel coronavirus, SARS-CoV-2, leading to severe pneumonia and acute respiratory disease, was observed in Wuhan, China (Figure 1) [1-4]. On January 21, 2020, the first confirmed case of COVID-19 was recorded in the United States [5]. As cases began to spread globally at an alarming rate, the World Health Organization officially recognized COVID-19 as a pandemic on March 11, 2020 [6]. Around the world, countries quickly implemented public health policies to mitigate the health and economic impacts caused by the virus [7]. The US federal government did not create a national unified pandemic control strategy [8-10], leading local state and city officials to implement their own public health and safety measures to "flatten the curve" [11]. However, by late May, many state and city leaders lifted their public health measures, leading to another alarming rise in COVID-19 case numbers [12,13]. As of October 29, 2020, there are 8,937,926 confirmed COVID-19 cases and 228,439 COVID-19–related deaths across the United States [14].

Figure 1. Timeline illustrating important events of the COVID-19 pandemic as related to US metropolitan areas.



Metropolitan areas largely drove the first wave of COVID-19 in the country [15], with northern travel hubs such as New York City, Chicago, Boston, Detroit, and Seattle seeing early spikes in the number of cases and deaths [14] (we refer to the area by the first city in its Census Bureau designation; eg, New York City for the New York-Newark-Jersey City, NY-NJ-PA core-based statistical area). Initial outbreaks around major east and west travel hubs are particularly important because they are gateways of international travel and have high population densities [16,17]. Genomic evidence from early SARS-CoV-2 strains suggests that initial infections in the United States were introduced to Washington State, a hub for travel to East Asia, by travelers from China [18-20]. There is also evidence that suggests New York City likely served as a primary entry point to the United States for the SARS-CoV-2 contagion from European hotspots in Spain and Italy early in the pandemic [19,21,22]. Chicago, Boston, and Detroit were likely additional entry points because of the large number of international travelers—Detroit's metro airport had over 1100 flights per day to 4 continents including direct flights to China and Europe [23], and was the least busy of the airports.

COVID-19 became endemic in the United States at the end of March [9], when many of these northern metro areas and travel hubs saw significant uptrends in their COVID-19 caseloads. In many of these areas, the first wave peaked in late March or early April after lockdowns and other preventive measures were imposed. For example, Washington State recorded its first death on February 19, 2020, and Governor Jay Inslee declared a state of emergency later that same day [24]. This led to increasingly stringent lockdown measures from Inslee, Seattle Mayor Jenny Durkan, and country executives, culminating in the governor issuing a "Stay Home Stay Safe" lockdown on March 23 [25,26]. Michigan's first case was recorded on March 10, and Michigan Governor Gretchen Whitmer imposed a "stay at

home" lockdown on March 23 [27,28]. In the Seattle metropolitan area, the first wave peaked the week of March 20-26, as it did in Detroit (authors' calculations using data from USAFacts.org [29]). New York peaked the week of March 27 to April 2, Boston the week of April 10-16, and Chicago the week of April 17-24. The ebb from the first peak is associated with the imposition of varying degrees of lockdown measures, social distancing, and mask wearing, among other COVID-19 responses and policies [11,30-33].

The first wave of the pandemic crested in southern cities about 3 months later, corresponding to a US-wide peak on July 16 with 77,352 new cases that day [19]. Phoenix peaked the week of July 3-9, with an average of 1404 new cases per day. Los Angeles peaked the week of July 10-16 with an average of 2080 new cases per day during that week; Miami also peaked that same week with an average of 1666 new cases per day.

The impact of the first wave on southern cities is associated with a lack of COVID-19 prevention measures in these areas [9,16,17,34,35]. For example, in Texas, on April 27, 2020, Governor Greg Abbott issued an executive order reopening the state including in-person retail and dining and prohibiting municipalities from imposing sanctions on individuals who chose not to wear a face mask [36], despite opposition from local government leaders [37,38]. On June 17, the governor allowed local jurisdictions to require face masks and, on July 2, mandated face masks across the state [37,38]. In Dallas and Houston, the pandemic initially peaked the week of July 10-16 and then ebbed (authors' calculations based on data from USAFacts.org [29]).

The United States is now entering the second wave of the COVID-19 pandemic [39,40]. As of November 2020, only 3 US counties have not reported a case of COVID-19, each of which has a population of less than 1000 people [41]. Many cities and states have reopened, citing the need to keep their economies functioning [42]. There has been considerable pushback against COVID-19 precautions, such as social distancing and face mask wearing. The pushbacks have reached the extremes of violent protests in Texas [43], and an alleged plot to kidnap the Governor of Michigan [44]. With the second—historically more deadly [45,46]—year of the pandemic coming up in a few months, surveillance of the disease will be even more relevant for determining appropriate COVID-19 precautions. The onset of the second wave requires taking initiative rather than relying on reactive public health measures, and in order to be proactive, an improved surveillance system is needed.

A robust surveillance system should answer relevant questions about the second wave: how many new cases appear in the metropolitan area per day per 100,000 population? Is the number of new cases an acceleration from the previous week, and is the acceleration indicative of explosive growth? Is there evidence of sustained transmission from new cases last week to new cases this week to new cases next week (eg, from Halloween, sports events, political rallies, etc)?

The objective of this paper is to provide novel, policy-relevant surveillance information about the second wave of COVID-19 cases in the 25 largest metropolitan areas in the United States. This information is of critical importance to controlling the second wave because in the absence of a national coronavirus policy, a COVID-19 policy is made and implemented at the state and local levels, including the metropolitan area [9,47]. Guidelines on reopening such as a 14-day downward trend in the number of new cases per day are difficult to implement at the local level without accurate quantitative data on the number of new cases and a local, area-specific trend. For example, in Los Angeles and New York City, day-to-day fluctuations in new case numbers were directly correlated to testing numbers rather than biological factors [48]. Additionally, there can be differences between state and local policies, leading to friction between governors and mayors with perhaps negative impacts on public adherence to social distancing, face mask wearing, and other public health guidelines and behaviors [49]. Consequently, surveillance information is needed at the metropolitan level to help inform policy needs and policy effectiveness.

Methods

Overview

This paper uses the dynamic panel data (DPD) modeling application and surveillance approach [50,51], which has been applied to sub-Saharan Africa, South Asia, and at the state level in the United States [52-54]. We apply these methods to the 25 most populous US metropolitan areas, as defined by the US Census Bureau [55]. For each metropolitan area, a panel data set is constructed using counties as the cross-sectional variable and days as the time variable.

The DPD modeling method is a novel approach to medical surveillance applications. Traditional contagion modeling techniques, including agent-based modeling and system dynamics, are complex, require sophisticated software (system dynamics), and are labor intensive, which makes them impractical for rapid surveillance across dozens of cities and counties in the United States, especially given the data that are readily available and easily accessible. Surveillance methods should rapidly generate understandable indicators to inform current decision making [56]. Our surveillance method can quickly be applied to existing data to generate indicators of pandemic outbreak locations and where explosive growth is likely to occur. The DPD-based modeling and surveillance method was validated previously [50,51]; these papers contain more detailed explanations and additional references to other methods.

Data

Data on the cumulative number of positive COVID-19 cases based on testing for each county and each day were downloaded from USAFacts.org [29] on October 17, 2020, as an excel file and are complete from January 22 through October 15, 2020. Data from March 20 to October 15 were used for statistical analysis and tables presented in this paper, which provides 30 weeks of data from the approximate start of the pandemic in the United States. On March 20, the country had 6367 new cases, the first day in which the number of new cases exceeded 5000 [57]. The only cleaning or processing of the data was to

create the novel surveillance metrics described below based on Oehmke et al [50].

Surveillance Methods

Following the procedures described by Oehmke et al [50], we calculated the novel surveillance metrics of speed, acceleration, and jerk for the COVID-19 caseloads for each of the 25 most populous metropolitan areas. Speed is defined as the number of new cases per day per 100,000 population; since reporting coverage depends on the day of the week, we report a weekly average value. Acceleration is the change in speed from one week to the next. It provides the primary indication of whether the pandemic is getting worse (positive acceleration) or better (negative acceleration or deceleration). Jerk is calculated as the change in acceleration from the prior week to the current week. Jerk is the second indictor of whether the pandemic is getting worse, with a positive jerk signaling growing acceleration that is possibly associated with a super-spreading event, ineffective reopening policy, colder weather, or other behavior or environmental change. A negative jerk indicates a slowing acceleration, possibly leading to a plateau or even a peak followed by deceleration in the pandemic.

Dynamic Panel Data Regression Methods

The DPD model generates the 7-day persistence surveillance indicator. This model relies in part on discerning common trends across counties within a metropolitan area, if present, to inform the modeling of overall metropolitan area trends in general and the 7-day persistence effect in particular.

We cannot replicate the full model described elsewhere [50,51] since testing data are available only on the number of positive tests administered by each county. Instead, we use the abbreviated model:

where the subscripts i and t denote the county within the metro area and the day of the measurement, respectively. $Speed_{it}$ is the number of new cases in county i on day t, ε_{it} is an error term, and the β_j parameters are those to be estimated. β_1 and β_2 quantify the 1-day and 2-day lag effects, and β_3 determines the base coefficient value for calculating the 7-day persistence effect. β_3 measures the number of new cases today that are statistically attributable to new cases 7 days ago, that is, it measures the propensity of the pandemic to travel in week-long "echoes" in which people newly diagnosed a week ago also infected others a week ago, and these others are diagnosed as new cases today. These week-long echoes could be caused by idiosyncratic factors such as super-spreading events and/or by systemic factors such as a systemic disregard for social distancing and mask wearing. The indicator variables $I_{10.9\text{-}10.15}$ and $I_{10.2\text{-}10.8}$ define the weeks October 9-15 and October 2-8, respectively, so that the coefficients β_4 and β_5 quantify weekly shifts in the 7-day persistence effect. A positive weekly shift could be caused by a super-spreading event that occurred during the week in question, by reopening, by the removal of mask-wearing requirements, or similar events.

We applied the DPD model to the 25 most populous metropolitan areas in the United States. The model was estimated using the generalized method of moments (GMM) approach [58] as implemented by Arellano and Bond [59] for DPD models and applied to the COVID-19 pandemic by Oehmke et al [50,51]. The Wald chi-square test was administered to test model fit based on the null hypothesis that the regression contains no explanatory power. The Sargan chi-square test was applied to determine model validity by testing the null hypothesis that the model is valid [58]. Statistical significance was determined at the 5% level. All estimations were conducted in STATA/MP 16.1 (StataCorp LLC) using the xtabond command.

Because of the use of lagged values in these formulae, we reported results for each of the 55 "weeks" (7-day periods) from Monday, March 2, 2020, through Sunday, January 3, 2021, in the Multimedia Appendices 1 and 2. This paper contains the results for the week of October 9-15, 2020.

Results

Persistence Rates

Complete data for persistence rates by International Standards Organization (ISO) week and metropolitan area are provided in an excel sheet (Multimedia Appendix 1). The full set of persistence results cover the week beginning on April 6, 2020, through the week ending on January 3, 2021. To optimize computer resources, weekly persistence rates were estimated using data for the annual quarter containing the week; for example, persistence rates for ISO week 53 (December 28, 2020, to January 3, 2021) are estimated from data for the fourth quarter of 2020.

We were unable to estimate a persistence rate for Phoenix for the week ending January 3, 2021, due to insufficient data at the time of estimation. The metropolitan areas of Los Angeles and Riverside comprise only 2 counties, and San Diego comprises 1 county, so at times there was insufficient cross-county variation for the application of DPD techniques. In this case, we estimated values for the combined Southern California region (combined Los Angeles, Riverside, and San Diego metro areas). In particular, persistence rates reported for Riverside and San Diego for the fourth quarter (ISO weeks 41 through 53) are estimates for the Southern California region.

A positive persistence rate most likely indicates continuing unsafe COVID-19 behaviors that recur over time, whereby infected individuals in a given week transmit the infection to other individuals who appear as COVID-19 cases the next week, leading to a "persistence" in the number of COVID-19 cases reported each week. Large positive persistence rates are associated with increasing case rates, and rates greater than 1 are indicators of potentially explosive growth. A negative persistence rate could indicate a choppy, up-and-down movement in the number of COVID-19 cases from week to week, or a period of downturn and decline in the number of cases.

Entering 2021, the metropolitan areas with the largest persistence were New York City (1.83), Miami (1.00),

Philadelphia (1.61), Tampa (1.08), Charlotte (1.29), and Orlando (1.04). These areas are at high risk for increased COVID-19 caseloads during the first full week of January, with a potential for explosive growth.

Surveillance Results

We report novel surveillance results of speed, acceleration, and jerk for each of the 44 "weeks" (7-day periods) from the week beginning on March 2, 2020, through the week ending on January 3, 2021, in Multimedia Appendix 2. Table 1 contains the results for the week of October 9-15, 2020. In relation to the timeline presented in Figure 1, notable findings are that for New York City and Seattle jerk turned negative the week of March 27 to April 2, 2020, 1 week after shutdowns. In Detroit, jerk turned negative the week of April 3-9, 2020, 2 weeks after the shutdown in Michigan. The negative jerks indicate a slowing of the pandemic acceleration; the chronological propinquity to the shutdown orders was consistent with a strong and rapid impact of these orders on the pandemic. The average number of daily new positive results for the week of October 9-15 ranged from 19 for San Antonio to 627 for Los Angeles (Table 2). The metropolitan areas with the greatest speed were Minneapolis (30) and Chicago (21). The metropolitan areas with the slowest speed were Portland (3) and San Francisco (5).

In the New York metropolitan area, the first wave peaked during the week of March 20-26, 2020, with an average of 12,855 new cases per day or a speed of over 68 new cases per day per 100,000 population. During the week of April 10-16, 2020, this area started to gain control of the pandemic, characterized by a negative acceleration (–10) and the area's largest negative jerk (–33). The next 8 weeks were characterized by negative acceleration, with speed declining to an average of 3 new cases per day per 100,000 population during the week of May 29 to June 4, 2020, and remaining at values of 3 or 4 before starting a second wave ascent the week of September 18-24, 2020. This general pattern was replicated to a large degree in other northern cities hit early by the first wave.

Although Dallas recorded its first 3 cases on March 9, case counts increased relatively slowly during an early Dallas shutdown to an average of 175 new cases daily and a speed of 2.3 new cases per 100,000 during the week of April 17-23. After the governor ordered Texas to reopen on April 17, case counts increased noticeably but not explosively, reaching a speed of 4.9 new daily cases per 100,000 the week of May 29 to June 4. During that week, acceleration and jerk turned positive and stayed positive through the week of July 10-16, possibly associated with the hot weather that encouraged people to visit crowded beaches or otherwise break social distancing and other COVID-19 protocols [60-62], which is considered risky behavior [63]. The first wave peaked in Dallas during the week of July 10-16 with an average of 2084 new cases per day that week and a speed of over 27.5 new cases per 100,000. Following the reimposition of COVID-19 prevention measures on July 2, acceleration turned negative and the number of cases declined during the weeks spanning July 17 to September 10, with the exception being the week of August 14-20 when a spike was associated with clearing a backlog of unreported prior test results [64]. However, even the trough during the week of September 4-10 had an average of 640 new cases per day and a speed of 8.4 new daily cases per 100,000, which is higher than during the early shutdown. In Dallas, the pandemic has now re-emerged with an average of 1116 new cases per day and a speed of 14.7 new daily cases per 100,000. Other southern metropolitan areas including Houston, Miami, Phoenix, and Tampa experienced similar first wave patterns, although evidence of a second wave has not hit all southern cities.

Table 1. Novel surveillance metrics for the week of October 9-15, 2020.

Metropolitan area	Number of daily new positive results, weekly average	Speed[a]	Acceleration[b]	Jerk[c]	7-day persistence effect[d]
1. New York	67	8	–1	–3	0.68
2. Los Angeles	627	8	0	0	2.90
3. Chicago	142	21	6	3	7.18
4. Dallas	101	11	1	1	1.54
5. Houston	73	8	–2	–5	2.47
6. Washington, DC	21	8	1	0	1.52
7. Miami	265	12	2	1	4.58
8. Philadelphia	65	9	2	2	1.48
9. Atlanta	26	12	1	1	1.14
10. Phoenix	255	10	2	0	2.78
11. Boston	70	9	2	1	2.16
12. San Francisco	42	5	–1	0	0.11
13. Riverside	263	11	0	0	3.51
14. Detroit	82	12	4	4	8.37
15. Seattle	94	7	1	1	5.21
16. Minneapolis	63	30	12	10	14.60
17. San Diego	286	9	0	0	2.31
18. Tampa	89	10	1	1	2.87
19. Denver	54	13	3	1	2.87
20. St Louis	37	22	5	12	3.18
21. Baltimore	38	8	0	0	2.29
22. Charlotte	40	18	0	4	3.72
23. Orlando	72	10	0	–2	2.81
24. San Antonio	19	3	–6	–7	–1.96
25. Portland	29	6	0	1	0.31

[a]Number of daily new positive results, weekly average.

[b]Change in speed between the weeks of October 2-8 and October 9-15.

[c]Change in acceleration between the weeks of October 2-8 and October 9-15.

[d]Number of cases this week statistically attributable to cases last week.

Table 2. Dynamic panel data regression of COVID-19 speed and 7-day persistence effect.

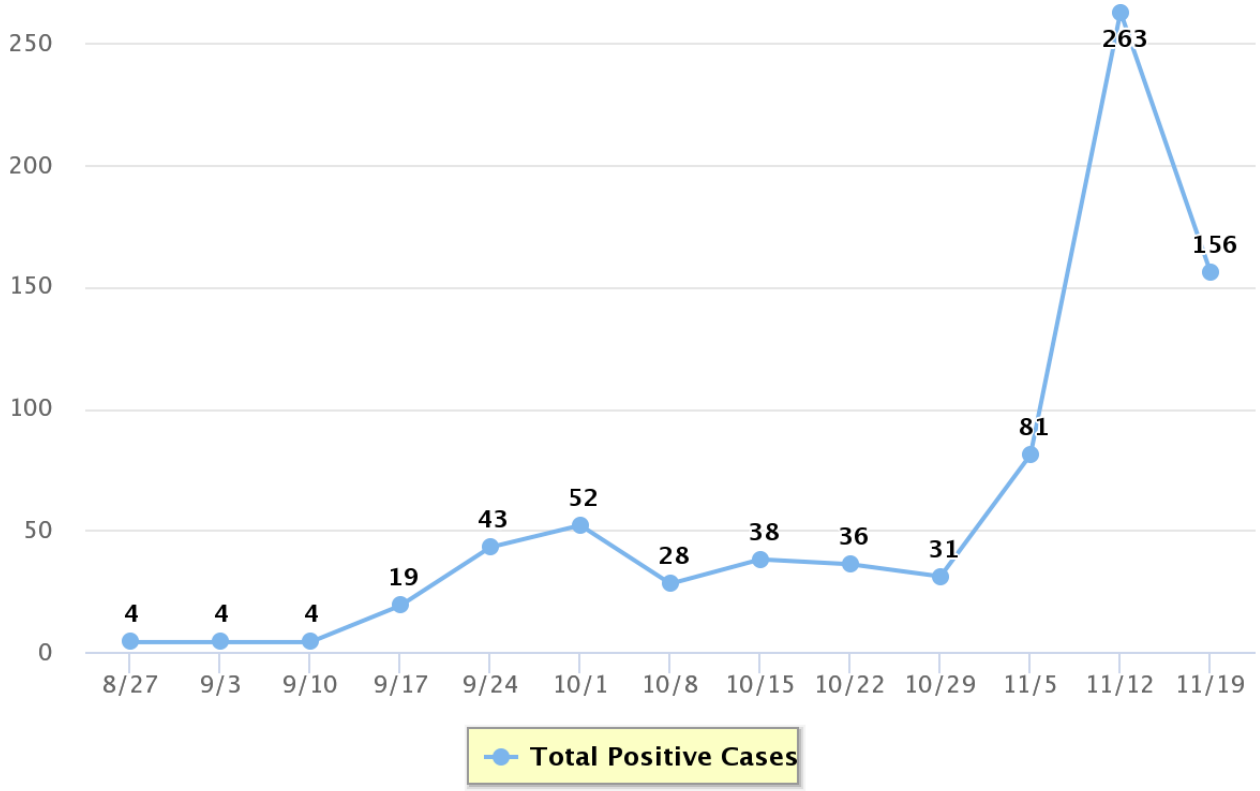
Metropolitan area (number of counties)	Wald χ^2 test of regression significance (*P* value)	Sargan χ^2 test of model validity (*P* value)	7-day persistence coefficient		
			Base effect (*P* value)	Shift, October 9-15 (*P* value)	Shift, October 2-8 (*P* value)
1. New York (n=23)	193.68 (<.001)	636.36 (.46)	0.0536 (.50)	0.0220 (.74)	0.2035 (.002)
2. Los Angeles (n=2)	25.03 (<.001)	49.87 (.60)	0.3101 (<.001)	0.0505 (.47)	0.0459 (.53)
3. Chicago (n=14)	130.28 (<.001)	418.07 (.27)	–0.0271 (.33)	0.5085 (<.001)	0.2238 (<.001)
4. Dallas (n=11)	48.67 (<.001)	322.71 (.36)	0.4104 (<.001)	–0.2602 (.003)	–0.3244 (.001)
5. Houston (n=9)	8.79 (.12)	257.55 (.46)	0.0951 (.08)	0.1389 (.36)	0.3367 (.09)
6. Washington, DC (n=25)	20.58 (.001)	736.88 (.11)	0.0250 (.55)	0.1931 (<.001)	0.0104 (.86)
7. Miami (n=3)	16.63 (.005)	79.46 (.56)	0.3279 (.04)	0.1088 (.35)	–0.0096 (.93)
8. Philadelphia (n=8)	8.32 (.14)	227.22 (.48)	0.0609 (.46)	0.1379 (.10)	–0.0097 (.90)
9. Atlanta (n=28)	46.85 (<.001)	851.42 (.14)	0.0929 (.02)	0.0117 (.80)	–0.0144 (.75)
10. Phoenix (n=2)	8.22 (.15)	50.58 (.57)	0.0454 (.73)	0.3060 (.07)	0.2838 (.12)
11. Boston (n=7)	27.97 (<.001)	196.40 (.52)	0.0973 (.29)	0.1853 (.04)	0.0172 (.85)
12. San Francisco (n=5)	10.59 (.06)	139.02 (.51)	0.1751 (.054)	–0.1543 (.20)	–0.2584 (.09)
13. Riverside (n=2)	34.37 (<.001)	47.07 (.70)	0.6413 (<.001)	–0.3360 (.06)	–0.1599 (.35)
14. Detroit (n=6)	245.98 (<.001)	189.22 (.14)	0.5095 (.37)	0.6095 (.45)	0.1853 (.03)
15. Seattle (n=3)	24.02 (<.001)	85.39 (.38)	0.1948 (.45)	0.6853 (.001)	0.4182 (.03)
16. Minneapolis (n=15)	115.52 (<.001)	531.86 (<.001)	0.0300 (.55)	0.8143 (<.001)	0.1926 (.06)
17. San Diego (n=1)	2.58 (.77)	24.75 (.42)	0.1719 (.29)	0.1003 (.45)	0.1227 (.36)
18. Tampa (n=4)	7.18 (.21)	113.02 (.43)	0.3055 (.02)	0.0307 (.77)	–0.0367 (.74)
19. Denver (n=10)	47.41 (<.001)	268.99 (.74)	–0.0811 (.48)	0.3725 (.001)	0.3194 (.003)
20. St Louis (n=15)	7.40 (.19)	443.33 (.32)	0.1025 (.13)	0.0855 (.37)	–0.1239 (.08)
21. Baltimore (n=7)	37.22 (<.001)	212.36 (.23)	0.1081 (.10)	0.1774 (.01)	0.2260 (.002)
22. Charlotte (n=11)	6.42 (.27)	314.04 (.49)	0.1578 (.08)	0.0525 (.59)	–0.0748 (.39)
23. Orlando (n=4)	20.00 (.001)	111.72 (.46)	0.1970 (.08)	0.0747 (.37)	0.1218 (.17)
24. San Antonio (n=8)	1.59 (.90)	224.59 (.53)	–0.0396 (.56)	–0.1654 (.75)	–0.1714 (.79)
25. Portland (n=7)	2.53 (.77)	197.38 (.50)	–0.2356 (.20)	0.2852 (.17)	0.2157 (.23)

Dynamic Panel Data Model Results

Table 1 contains the regression diagnostics for the models. The Sargan tests of model validity failed to reject the null hypothesis of validity for all metro areas except for Minneapolis (χ^2_5=531.86, *P*<.001), showing model validity for all areas but Minneapolis. Rejection of the null hypothesis for the Sargan test for Minneapolis means that the model is not valid for this metro area and is insufficient to capture the rapid acceleration and jerk exhibited by Minneapolis. Minneapolis exhibited extreme behavior, with its speed jumping from an average number of 17 daily new positive results standardized per 100,000 population for the week of October 2-8 to 30 (67% increase) for the week of October 9-15. No other metropolitan area has exhibited that high a speed for the week of October 9-15.

Potential risk factors for the Minneapolis increase include the advent of colder fall weather, although the week of October 9-15 was warmer than average [65], and the phase II reopening of the University of Minnesota. During the prior weeks, the number of new cases reported at the University of Minnesota increased from 4 in the beginning of September to 52 the first week in October (Figure 2), which numerically contributes to Minneapolis' rapid acceleration and large jerk. The influence of universities reopening is consistent with state-level findings reported by Oehmke et al [50].

Figure 2. Number of positive COVID-19 cases recorded by the University of Minnesota, Twin Cities.



The Wald test of model significance shows that the model explains a statistically significant proportion of the variation in the caseload for 16 of the 25 most populous metropolitan areas. The lack of statistical significance in the other 9 most populous metropolitan areas is indicative of the absence of both 1-day and 7-day persistence effects, since the regression equation is designed specifically to measure these persistence effects. The lack of a persistence effect is most likely to occur when the number of new cases per day is relatively flat (constant speed). For example, the Wald test for the Tampa metropolitan area is 7.18 (*P*=.21), and for the weeks included in the regression analysis the speed was 9, 10, 8, 9, and 10, respectively. In other words, the lack of a significant persistence effect is consistent with and corroborates the finding of a low-level and flat profile for the pandemic in the Tampa metropolitan area during those weeks. Of the 9 areas where the model was not statistically significant, 8 had fewer than 10 counties. In the ninth area, St Louis, the speed ranged from 17 to 24 for the past 3 months with no discernable upward or downward trend—in this case, a simple constant provides a reasonably good model and the DPD model contributes only minimally.

Detroit's speed increased from 59 for the week October 2-8 to 82 in the week October 9-15. Its persistence factor increased by 0.6095 for the week of October 9-15 from 0.5095 to 1.290, which is indicative of explosive growth, although neither the base persistence effect nor the weekly shift effect was individually statistically significant. Chicago had one of the largest percentage increases in speed, from 87 the week of September 25 to October 1 to 100 for the week of October 2-8 to 142 for the week of October 9-15, an 80% increase in just 2 weeks, and Chicago's persistence for the weeks of October 2-8 and October 9-15 increased by 0.2238 and 0.5085, respectively. Minneapolis had the second highest increase in speed over the 2 weeks, from 38 at the beginning of the 2-week period to 42 at the end of the second week to 63 at the end of the third week. This represents a 67% increase in just 14 days. For the week of October 9-15, Minneapolis experienced a persistence shift of 0.8143, the largest persistence shift recorded during the 2-week period. An increasing persistence (positive and statistically significant shift) would be expected to lead to greater speed in the next week.

Discussion

Principal Findings

The DPD model successfully represented COVID-19 case dynamics for all metropolitan areas, except for the Minneapolis-St. Paul-Bloomington area that experienced extreme events. Calculating speed, acceleration, and jerk also helped with understanding case dynamics in metro areas, especially as they refer to characterizing a peak or wave progression.

The first wave of COVID-19 hit northern metropolitan areas first. These areas were able to flatten the curve by imposing shutdowns, social distancing, mask wearing, and other COVID-19 protocols. The first wave did not hit southern metropolitan areas as early or as hard as northern areas; southern metropolises reopened and only after reopening did the first wave truly hit. The current situation in the New York metropolitan area exemplifies the usefulness of the novel

metrics. The second wave hit New York during the week of August 28 to September 3, based on changes in acceleration and jerk from negative to positive values. The onset was confirmed the week of September 11-17, when acceleration nearly tripled from the prior week, and jerk more than quadrupled to a value of 0.98. However, the governors of New York and New Jersey did not impose additional restrictions until November 11, and then the restrictions were primarily curfews on bars and late-night entertainment, based on information available at the time [66]. Use of novel metrics could have influenced earlier, more forceful, and more proactive COVID-19 policy.

Limitations

A limitation of this data set is that it does not have data on the total number of tests reported per day at the county level. A limitation of the estimation technique is that the metropolitan areas of San Diego and Riverside, comprising one and two counties, respectively, had insufficient cross-sectional information and the estimation had to be conducted for a combined southern California region.

Conclusion

Our analysis including the use of novel surveillance metrics shows that the second wave of the pandemic has arrived in the United States and is accelerating, especially in northern metropolitan areas. For metropolitan areas in the Midwest, specifically Chicago, Detroit, and Minneapolis, there has been rapid and potentially explosive growth in cases during the first half of October 2020. This type of growth can be seen from the cities' speed, acceleration, and jerk, as well as the increasing 7-day persistence effects. It is critical for those cities already feeling the second wave to react swiftly and strongly. For those cities who have so far escaped the second wave, it is critically important to studiously monitor surveillance data to ascertain if and when the second wave is beginning to hit, and then to be proactive in reimposing COVID-19 protocols. The overall conclusion is that improved COVID-19 surveillance metrics can help cities be proactive in managing the pandemic, leading to fewer cases and saving lives.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Seven-day persistence by week and metropolitan area, 2020.
[XLSX File (Microsoft Excel File), 21 KB - jmir_v23i2e26081_app1.xlsx]

Multimedia Appendix 2
Speed, acceleration, and jerk by metropolitan area and week, 2020.
[XLSX File (Microsoft Excel File), 74 KB - jmir_v23i2e26081_app2.xlsx]

References

1. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet 2020 Feb;395(10223):507-513. [doi: 10.1016/s0140-6736(20)30211-7]
2. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207. [doi: 10.1056/nejmoa2001316]
3. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017]
4. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
5. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, Washington State 2019-nCoV Case Investigation Team. First Case of 2019 Novel Coronavirus in the United States. N Engl J Med 2020 Mar 05;382(10):929-936 [FREE Full text] [doi: 10.1056/NEJMoa2001191] [Medline: 32004427]
6. Mahase E. Covid-19: WHO declares pandemic because of "alarming levels" of spread, severity, and inaction. BMJ 2020 Mar 12;368:m1036. [doi: 10.1136/bmj.m1036] [Medline: 32165426]
7. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? The Lancet 2020 Mar;395(10228):931-934. [doi: 10.1016/s0140-6736(20)30567-5]
8. Maor M, Howlett M. Explaining variations in state COVID-19 responses: psychological, institutional, and strategic factors in governance and public policy-making. Policy Design and Practice 2020 Sep 25;3(3):228-241. [doi: 10.1080/25741292.2020.1824379]
9. Haffajee RL, Mello MM. Thinking Globally, Acting Locally — The U.S. Response to Covid-19. N Engl J Med 2020 May 28;382(22):e75. [doi: 10.1056/nejmp2006740]

10. Kettl DF. States Divided: The Implications of American Federalism for Covid-19. Public Adm Rev 2020 May 22;80(4):595-602 [FREE Full text] [doi: 10.1111/puar.13243] [Medline: 32836439]
11. Laestadius L, Wang Y, Ben Taleb Z, Kalan M, Cho Y, Manganello J. Online National Health Agency Mask Guidance for the Public in Light of COVID-19: Content Analysis. JMIR Public Health Surveill 2020 May 26;6(2):e19501 [FREE Full text] [doi: 10.2196/19501] [Medline: 32427103]
12. Aleta A, Martín-Corral D, Pastore Y Piontti A, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nat Hum Behav 2020 Sep 05;4(9):964-971. [doi: 10.1038/s41562-020-0931-9] [Medline: 32759985]
13. Ngonghala CN, Iboi EA, Gumel AB. Could masks curtail the post-lockdown resurgence of COVID-19 in the US? Math Biosci 2020 Nov;329:108452 [FREE Full text] [doi: 10.1016/j.mbs.2020.108452] [Medline: 32818515]
14. CDC COVID Data Tracker Internet. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/covid-data-tracker/#cases [accessed 2020-10-31]
15. Rader B, Scarpino SV, Nande A, Hill AL, Adlam B, Reiner RC, et al. Crowding and the shape of COVID-19 epidemics. Nat Med 2020 Dec 05;26(12):1829-1834. [doi: 10.1038/s41591-020-1104-0] [Medline: 33020651]
16. Chin T, Kahn R, Li R, Chen J, Krieger N, Buckee C, et al. US-county level variation in intersecting individual, household and community characteristics relevant to COVID-19 and planning an equitable response: a cross-sectional analysis. BMJ Open 2020 Sep 01;10(9):e039886 [FREE Full text] [doi: 10.1136/bmjopen-2020-039886] [Medline: 32873684]
17. Browne A, Ahmad SSO, Beck C, Nguyen-Van-Tam J. The roles of transportation and transportation hubs in the propagation of influenza and coronaviruses: a systematic review. J Travel Med 2016 Jan;23(1):a [FREE Full text] [doi: 10.1093/jtm/tav002] [Medline: 26782122]
18. CDC COVID-19 Response Team, Jorden MA, Rudman SL, Villarino E, Hoferka S, Patel MT, et al. Evidence for Limited Early Spread of COVID-19 Within the United States, January-February 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 05;69(22):680-684 [FREE Full text] [doi: 10.15585/mmwr.mm6922e1] [Medline: 32497028]
19. Schuchat A, CDC COVID-19 Response Team. Public Health Response to the Initiation and Spread of Pandemic COVID-19 in the United States, February 24-April 21, 2020. MMWR Morb Mortal Wkly Rep 2020 May 08;69(18):551-556 [FREE Full text] [doi: 10.15585/mmwr.mm6918e2] [Medline: 32379733]
20. Fauver JR, Petrone ME, Hodcroft EB, Shioda K, Ehrlich HY, Watts AG, et al. Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States. Cell 2020 May 28;181(5):990-996.e5 [FREE Full text] [doi: 10.1016/j.cell.2020.04.021] [Medline: 32386545]
21. Bushman D, Alroy KA, Greene SK, Keating P, Wahnich A, Weiss D, CDC COVID-19 Surge Laboratory Group;, et al. Detection and Genetic Characterization of Community-Based SARS-CoV-2 Infections - New York City, March 2020. MMWR Morb Mortal Wkly Rep 2020 Jul 17;69(28):918-922 [FREE Full text] [doi: 10.15585/mmwr.mm6928a5] [Medline: 32678072]
22. Maurano MT, Ramaswami S, Zappile P, Dimartino D, Boytard L, Ribeiro-Dos-Santos AM, et al. Sequencing identifies multiple early introductions of SARS-CoV-2 to the New York City region. Genome Res 2020 Dec 22;30(12):1781-1788 [FREE Full text] [doi: 10.1101/gr.266676.120] [Medline: 33093069]
23. About DTW. Wayne County Airport Authority. 2020. URL: https://www.metroairport.com/about-dtw [accessed 2021-01-26]
24. Inslee Issues COVID-19 Emergency Proclamation. Office of the Washington Governor Jay Inslee. 2020 Feb 29. URL: https://www.governor.wa.gov/news-media/inslee-issues-covid-19-emergency-proclamation [accessed 2020-11-05]
25. Inslee announces “Stay Home, Stay Healthy” order. Office of the Washington Governor Jay Inslee. 2020 Mar 23. URL: https://www.governor.wa.gov/news-media/inslee-announces-stay-home-stay-healthy%C2%A0order [accessed 2020-11-05]
26. McNerthney C. Coronavirus in Washington state: A timeline of the outbreak through March 2020. KIRO 7 News. 2020 Apr 3. URL: https://www.kiro7.com/news/local/coronavirus-washington-state-timeline-outbreak/IM65JK66N5BYTIAPZ3FUZSKMUE/ [accessed 2021-01-26]
27. Executive Order 2020-21 (COVID-19) – Rescinded. Office of Governor Gretchen Whitmer. 2020. URL: https://www.michigan.gov/whitmer/0,9309,7-387-90499_90705-522626--,00.html [accessed 2021-01-26]
28. Haddad K. Michigan issues stay-at-home order amid coronavirus: Here’s what it means. Click On Detroit. Michigan issues stay-at-home order amid coronavirus: Here’s what it means. Click on Detroit; 2020 . URL: https://www.clickondetroit.com/news/2020/03/23/michigan-issues-stay-at-home-order-amid-coronavirus-heres-what-it-means/ [accessed 2021-01-26]
29. US Coronavirus Cases and Deaths. USA Facts. URL: https://usafacts.org/visualizations/coronavirus-covid-19-spread-map [accessed 2020-11-01]
30. Javid B, Weekes MP, Matheson NJ. Covid-19: should the public wear face masks? BMJ 2020 Apr 09;369:m1442. [doi: 10.1136/bmj.m1442] [Medline: 32273278]
31. Greenhalgh T, Schmid MB, Czypionka T, Bassler D, Gruer L. Face masks for the public during the covid-19 crisis. BMJ 2020 Apr 09;369:m1435. [doi: 10.1136/bmj.m1435] [Medline: 32273267]
32. Cheng KK, Lam TH, Leung CC. Wearing face masks in the community during the COVID-19 pandemic: altruism and solidarity. The Lancet 2020 Apr:1-2. [doi: 10.1016/S0140-6736(20)30918-1]

33. Dalton C, Corbett S, Katelaris A. Pre-Emptive Low Cost Social Distancing and Enhanced Hygiene Implemented before Local COVID-19 Transmission Could Decrease the Number and Severity of Cases. SSRN Journal 2020 Mar 05:1-8. [doi: 10.2139/ssrn.3549276]
34. Weiner Z, Wong G, Elbanna A, Tkachenko A, Maslov S, Goldenfeld N. Projections and early-warning signals of a second wave of the COVID-19 epidemic in Illinois. medRxiv. Preprint posted online July 7, 2020 2021. [doi: 10.1101/2020.07.06.20147868]
35. Liu M, Thomadsen R, Yao S. Forecasting the spread of COVID-19 under different reopening strategies. Sci Rep 2020 Nov 23;10(1):20367 [FREE Full text] [doi: 10.1038/s41598-020-77292-8] [Medline: 33230234]
36. Abbott G. Executive Order No GA-18 relating to the expanded reopening of services as part of the safe, strategic plan to Open Texas in response to the COVID- 19 disaster. Legislative Reference Library of Texas. URL: https://lrl.texas.gov/scanned/govdocs/Greg%20Abbott/2020/GA-18.pdf [accessed 2021-01-26]
37. Abbott G. Executive Order No GA-29 relating to the use of face coverings during the COVID- 19 disaster. Office of the Texas Governor. URL: https://gov.texas.gov/uploads/files/press/EO-GA-29-use-of-face-coverings-during-COVID-19-IMAGE-07-02-2020.pdf [accessed 2021-01-26]
38. Friend D. COVID-19 in Texas: A Timeline of Gov. Abbott's Executive Orders and Case Numbers. The Texan. 2020 Sep 07. URL: https://thetexan.news/covid-19-in-texas-a-timeline-of-gov-abbotts-executive-orders-and-case-numbers/ [accessed 2021-01-26]
39. Solis J, Franco-Paredes C, Henao-Martínez AF, Krsak M, Zimmer S. Structural Vulnerability in the U.S. Revealed in Three Waves of COVID-19. Am J Trop Med Hyg 2020 Jul;103(1):25-27 [FREE Full text] [doi: 10.4269/ajtmh.20-0391] [Medline: 32383432]
40. Taskinsoy J. Coronavirus Pandemic: The U.S. Induced Great Panic of the 21st Century. SSRN Journal 2020 May 26:1-15. [doi: 10.2139/ssrn.3610749]
41. Smith S. Loving county, Texas, is one of three U.S. counties with zero Covid-19 cases. CNBC. 2020 Oct 23. URL: https://www.cnbc.com/video/2020/10/23/loving-county-texas-is-one-of-three-counties-in-the-u-s-with-zero-covid-19-cases.html [accessed 2021-01-26]
42. Li H, Jecker NS, Chung RY. Reopening Economies during the COIVD-19 Pandemic: Reasoning about Value Tradeoffs. Am J Bioeth 2020 Jul 27;20(7):136-138. [doi: 10.1080/15265161.2020.1779406] [Medline: 32716790]
43. Fernandez M, Montgomery D. Businesses Chafing Under Covid-19 Lockdowns Turn to Armed Defiance. The New York Times. 2020 May 13. URL: https://www.nytimes.com/2020/05/13/us/coronavirus-businesses-lockdown-guns.html [accessed 2020-11-13]
44. Six Arrested on Federal Charge of Conspiracy to Kidnap the Governor of Michigan. The United Stated Department of Justice. 2020 Oct 8. URL: https://www.justice.gov/opa/pr/six-arrested-federal-charge-conspiracy-kidnap-governor-michigan [accessed 2021-01-26]
45. Akin L, Gözel MG. Understanding dynamics of pandemics. Turkish Journal of Medical Sciences 2020 Apr 21;50(SI-1):515-519 [FREE Full text] [doi: 10.3906/sag-2004-133] [Medline: 32299204]
46. Johnson NPAS, Mueller J. Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. Bull Hist Med 2002;76(1):105-115. [doi: 10.1353/bhm.2002.0022] [Medline: 11875246]
47. Government Response to Coronavirus, COVID-19. US Government. 2020. URL: https://www.usa.gov/coronavirus [accessed 2021-01-26]
48. Bergman A, Sella Y, Agre P, Casadevall A. Oscillations in U.S. COVID-19 Incidence and Mortality Data Reflect Diagnostic and Reporting Factors. American Society of Microbiology 2020 Jul 14;5(4):1-6 [FREE Full text] [doi: 10.1128/mSystems.00544-20] [Medline: 32665331]
49. Greenblatt A. Will State Preemption Leave Cities More Vulnerable? Governing: The Future of States and Localities. 2020 Apr 3. URL: https://www.governing.com/now/Will-State-Preemption-Leave-Cities-More-Vulnerable.html [accessed 2021-01-26]
50. Oehmke JF, Moss CB, Singh LN, Oehmke TB, Post LA. Dynamic Panel Surveillance of COVID-19 Transmission in the United States to Inform Health Policy: Observational Statistical Study. J Med Internet Res 2020 Oct 5;22(10):e21955 [FREE Full text] [doi: 10.2196/21955] [Medline: 32924962]
51. Oehmke J, Oehmke T, Singh L, Post L. Dynamic Panel Estimate-Based Health Surveillance of SARS-CoV-2 Infection Rates to Inform Public Health Policy: Model Development and Validation. J Med Internet Res 2020 Sep 22;22(9):e20924 [FREE Full text] [doi: 10.2196/20924] [Medline: 32915762]
52. Post L, Kulasekere D, Moss C, Murphy R, Achenbach C, Ison M, et al. South Asia SARS-CoV-2 Surveillance System: The Race to the Top Using Longitudinal Trend Analysis. JMIR Preprints. Preprint posted online Nov 13, 2020 2021. [doi: 10.2196/preprints.25753]
53. Post L, Argaw S, Jones C, Moss C, Resnick D, Singh L, et al. A SARS-CoV-2 Surveillance System in Sub-Saharan Africa: Modeling Study for Persistence and Transmission to Inform Policy. J Med Internet Res 2020 Nov 19;22(11):e24248 [FREE Full text] [doi: 10.2196/24248] [Medline: 33211026]

54. Post LA, Issa TZ, Boctor MJ, Moss CB, Murphy RL, Ison MG, et al. Dynamic Public Health Surveillance to Track and Mitigate the US COVID-19 Epidemic: Longitudinal Trend Analysis Study. J Med Internet Res 2020 Dec 03;22(12):e24286 [FREE Full text] [doi: 10.2196/24286] [Medline: 33216726]
55. Metropolitan and Micropolitan Statistical Areas Population Totals and Components of Change: 2010-2019. The United States Census Bureau. URL: https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-metro-and-micro-statistical-areas.html [accessed 2020-10-31]
56. Teutsch S, Thacker S. Planning a public health surveillance system. Epidemiol Bull 1995 Mar;16(1):1-6. [Medline: 7794696]
57. Home - Johns Hopkins Coronavirus Resource Center. Johns Hopkins University Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/ [accessed 2020-10-31]
58. Hansen LP. Large Sample Properties of Generalized Method of Moments Estimators. Econometrica 1982 Jul;50(4):1029. [doi: 10.2307/1912775]
59. Arellano M, Bond S. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. The Review of Economic Studies 1991 Apr;58(2):277-297. [doi: 10.2307/2297968]
60. Gaspar E, Goulding A, Seal E, Vonderhaar J. How Texans celebrated July Fourth during the coronavirus pandemic. The Texas Tribune. 2020 Jul 05. URL: https://www.texastribune.org/2020/07/05/july-4-texas-coronavirus/ [accessed 2021-01-26]
61. Montgomery D, Fernandez M. 44 Texas Students Have Coronavirus After Spring Break Trip. The New York Times. 2020 Apr 01. URL: https://www.nytimes.com/2020/04/01/us/coronavirus-texas-austin-spring-break-cabo.html [accessed 2020-11-13]
62. Mangrum D, Niekamp P. College Student Contribution to Local COVID-19 Spread. SSRN Journal. Preprint posted online May 21, 2020 2021 [FREE Full text] [doi: 10.2139/ssrn.3606811]
63. Visiting Beaches and Pools. Centers for Disease Controls and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/beaches-pools.html [accessed 2021-01-26]
64. Carter S. Dallas County Adds 5,000 New COVID-19 Cases to July's Count. Dallas Observer. 2020 Aug 18. URL: https://www.dallasobserver.com/news/dallas-county-counts-5000-additional-cases-thanks-to-covid-19-testing-backlog-11935823 [accessed 2021-01-26]
65. Minneapolis, MN Weather History. Weather Underground. URL: https://www.wunderground.com/history/monthly/us/mn/minneapolis [accessed 2020-11-01]
66. Millman J. Cuomo Imposes Curfew on NY Bars, Restaurants and Gyms, Limits Private Home Gatherings to 10. NBC New York. 2020 Nov 12. URL: https://www.nbcnewyork.com/news/coronavirus/new-ny-nj-cases-soar-nearly-50-in-a-week-u-s-could-hit-20-million-cases-by-christmas/2716714/ [accessed 2021-01-26]

Abbreviations

DPD: dynamic panel data
GMM: generalized method of moments
ISO: International Standards Organization

Edited by G Eysenbach; submitted 01.12.20; peer-reviewed by M Elbattah; comments to author 22.12.20; revised version received 12.01.21; accepted 20.01.21; published 09.02.21.

Please cite as:
Oehmke TB, Post LA, Moss CB, Issa TZ, Boctor MJ, Welch SB, Oehmke JF
Dynamic Panel Data Modeling and Surveillance of COVID-19 in Metropolitan Areas in the United States: Longitudinal Trend Analysis
J Med Internet Res 2021;23(2):e26081
URL: https://www.jmir.org/2021/2/e26081
doi:10.2196/26081
PMID:33481757

Original Paper

Understanding the Public Discussion About the Centers for Disease Control and Prevention During the COVID-19 Pandemic Using Twitter Data: Text Mining Analysis Study

Joanne Chen Lyu[1], PhD; Garving K Luli[2], PhD

[1]Center for Tobacco Control Research and Education, University of California, San Francisco, San Francisco, CA, United States

[2]Department of Mathematics, University of California, Davis, Davis, CA, United States

Corresponding Author:
Joanne Chen Lyu, PhD
Center for Tobacco Control Research and Education
University of California, San Francisco
530 Parnassus Avenue
San Francisco, CA, 94143-1390
United States
Phone: 1 415 502 4181
Email: chenjoanne.lyu@ucsf.edu

Abstract

Background: The Centers for Disease Control and Prevention (CDC) is a national public health protection agency in the United States. With the escalating impact of the COVID-19 pandemic on society in the United States and around the world, the CDC has become one of the focal points of public discussion.

Objective: This study aims to identify the topics and their overarching themes emerging from the public COVID-19-related discussion about the CDC on Twitter and to further provide insight into public's concerns, focus of attention, perception of the CDC's current performance, and expectations from the CDC.

Methods: Tweets were downloaded from a large-scale COVID-19 Twitter chatter data set from March 11, 2020, when the World Health Organization declared COVID-19 a pandemic, to August 14, 2020. We used R (The R Foundation) to clean the tweets and retain tweets that contained any of five specific keywords—cdc, CDC, centers for disease control and prevention, CDCgov, and cdcgov—while eliminating all 91 tweets posted by the CDC itself. The final data set included in the analysis consisted of 290,764 unique tweets from 152,314 different users. We used R to perform the latent Dirichlet allocation algorithm for topic modeling.

Results: The Twitter data generated 16 topics that the public linked to the CDC when they talked about COVID-19. Among the topics, the most discussed was COVID-19 death counts, accounting for 12.16% (n=35,347) of the total 290,764 tweets in the analysis, followed by general opinions about the credibility of the CDC and other authorities and the CDC's COVID-19 guidelines, with over 20,000 tweets for each. The 16 topics fell into four overarching themes: knowing the virus and the situation, policy and government actions, response guidelines, and general opinion about credibility.

Conclusions: Social media platforms, such as Twitter, provide valuable databases for public opinion. In a protracted pandemic, such as COVID-19, quickly and efficiently identifying the topics within the public discussion on Twitter would help public health agencies improve the next-round communication with the public.

(J Med Internet Res 2021;23(2):e25108) doi:10.2196/25108

KEYWORDS

COVID-19; CDC; Twitter; public discussion; public opinion; social media; data mining

Introduction

Since first identified in Wuhan, China, in December 2019, COVID-19 has spread rapidly around the world. On March 11, 2020, the World Health Organization (WHO) declared the coronavirus outbreak a pandemic and was unable to determine the duration of the pandemic [1]. As of August 5, 2020, 213 countries and territories around the world have reported a total of 18,939,540 confirmed cases and a death toll of 709,700 [2].

The United States reported its first case on January 20, 2020 [3]; the country had 4,802,491 total COVID-19 cases and 157,631 total deaths as of August 5, 2020 [4].

The Centers for Disease Control and Prevention (CDC) is the "nation's health protection agency, working 24/7 to protect America from health and safety threats" [5]. Since January 21, 2020, the CDC has launched an agencywide response to this pandemic, including preparing health care providers and health systems, supporting governments at various levels on the front lines, and learning and sharing COVID-19 knowledge via a variety of communication channels. Amid the unprecedented public health crisis caused by the pandemic, the bewildered public in dire need of guidance depends on the quick response of public health authorities and is in greater demand for information and advice from them [6]. Previous studies found that the public willingness to take the advice (eg, handwashing) proposed by public health agencies, which will further impact the success of disease control strategies and policies, is related to the trust that the public has in the agencies [7-12]. During the Zika outbreak, studies found a substantial topic discrepancy between public concern and the CDC's response to Zika [13-16], undermining the efficacy of the CDC's Zika control efforts. Considering that the COVID-19 pandemic will last for a protracted time, timely information about public opinion regarding the CDC's COVID-19 response efforts and their concerns about COVID-19 can provide insight to improve the next round of communication with the public.

Social media platforms such as Twitter have not only become increasingly important for the public to seek, share, and discuss information, but have also provided valuable platforms for the surveillance of public opinion, allowing for the monitoring of the public's questions, concerns, and responses to health threats [17-19]. In previous public crises, such as the Ebola virus [20], Zika virus [13,17,21], and H1N1 virus outbreaks [22], Twitter was used as an up-to-date information source to gauge the public's knowledge and reactions to the epidemics. During the Zika outbreak, it was found that inaccurate information proliferated on social media, and conspiracy theories regarding the Zika virus were more popular than public health education materials from health agencies [23]. Therefore, using social media to monitor public knowledge, to evaluate the information spread online, and to address the identified problems in a timely manner is crucial in battling public health crises. In addition, when an epidemic situation is not clear, the discussion on social media can provide timely information for improving epidemic surveillance and forecasting [24-27]. Therefore, the value of online discussion on social media platforms, especially during infectious disease epidemics, has gained ever more attention by public health agencies and officials [24-26]. The CDC has been actively using Twitter to reach out with timely health and safety information [28] and even hosted live Twitter chats to directly communicate with the general public in the periods of the Ebola outbreak [20] and the Zika outbreak [13].

The ongoing COVID-19 pandemic demands continuous and evolving efforts of using social media data to understand the public's thoughts and concerns. While a series of studies made significant contributions along these lines, the majority of existing studies analyzed data collected before the middle of April 2020 [29-39]. In addition, there have been scarce social media studies with a focus on the CDC during the COVID-19 pandemic [6], the main source for evidence-based information about the pandemic [40]. To fill the gaps in knowledge after the WHO's declaration of a pandemic, we used text mining methods to analyze COVID-19-related tweets about the CDC from March 11 to August 14, 2020. By doing so, this study identified the topics emerging from the tweets and the overarching themes of these topics, shedding light on a series of questions: What are the public concerns over COVID-19? What does the public expect from the CDC? and How does the public comment on the current performance of the CDC in response to COVID-19?

Methods

Data Extraction and Preprocessing

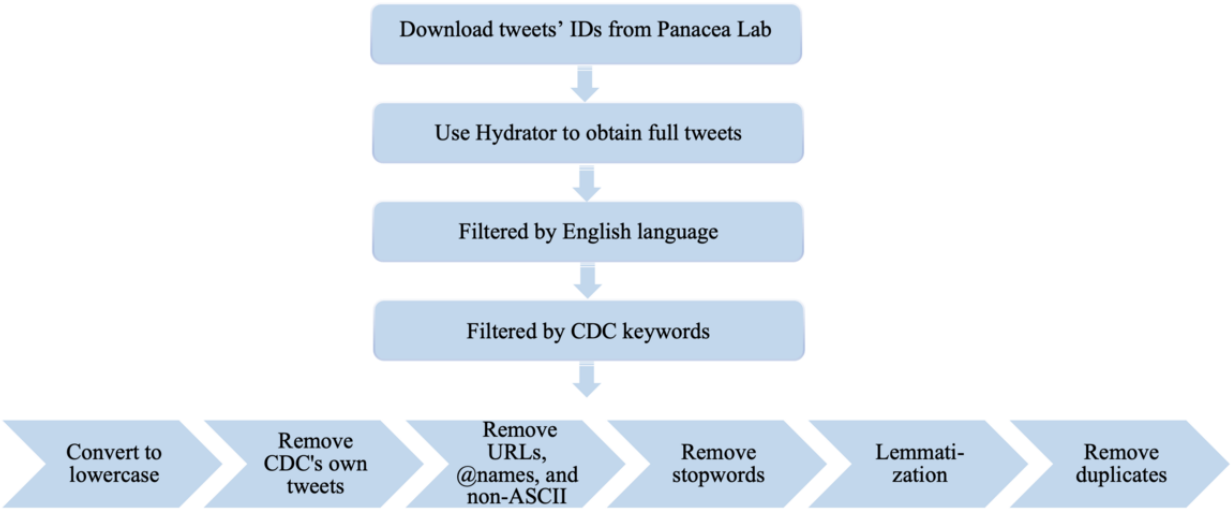
The IDs from a total of 128,432,021 tweets, without retweets, from March 11 through August 14, 2020, were obtained using the data set maintained by Georgia State University's Panacea Lab [41]. These tweets were collected by the Panacea Lab using the following 13 keywords: COVD19, CoronavirusPandemic, COVID-19, 2019nCoV, CoronaOutbreak, coronavirus , WuhanVirus, covid19, coronaviruspandemic, covid-19, 2019ncov, coronaoutbreak, and wuhanvirus. Since Twitter's Terms of Service do not allow the full JavaScript Object Notation (JSON) for data sets of tweets to be distributed to third parties, Georgia State University's Panacea Lab can only provide tweet IDs [42], which can be hydrated to obtain the JSON objects from these tweets.

During the tokenizing stage, we used the *gsub* function in R (The R Foundation) to extract the tweets whose language field in the tweets' metadata was specified as English. All text mining was done using R 4.0.2 GUI (graphical user interface) 1.72 Catalina build (7847) on a Mac running 10.15 Catalina.

We converted all the tweets to lowercase and created a script to remove the URLs, mentioned names, non-ASCII (American Standard Code for Information Interchange) characters such as emojis, and anything other than English letters or spaces (eg, "1," "?," etc). Using the R package *dplyr*, version 1.0.2, we cleaned the tweets by removing duplicates, retained only tweets that contained any of five specific keywords—cdc, CDC, centers for disease control and prevention, CDCgov, and cdcgov—and eliminated all 91 tweets posted by the CDC itself. The final cleaned data set consisted of 290,764 unique tweets from 152,314 different users. We further cleaned the tweets by removing words and characters that were of little or no analytical value (eg, "the," "very," "&," etc). We performed this task by creating our own list of stop words by appending the 13 keywords related to "COVID19" and the five keywords related to "CDC" to the English stop words list from the R package *tidytext*, version 0.2.6; this was done because we already knew that every tweet would contain one or more of those keywords, and having them in the tweets does not contribute to further our understanding of the main content of the tweets. Lastly, we stemmed and lemmatized the words to their root forms using the R package *textstem*, version 0.1.4 (eg, *studying*, *studies*, and

studied were converted to *study*). See Figure 1 for a summary of our data extraction and preprocessing procedure.

Figure 1. Flowchart for data extraction and preprocessing. CDC: Centers for Disease Control and Prevention.

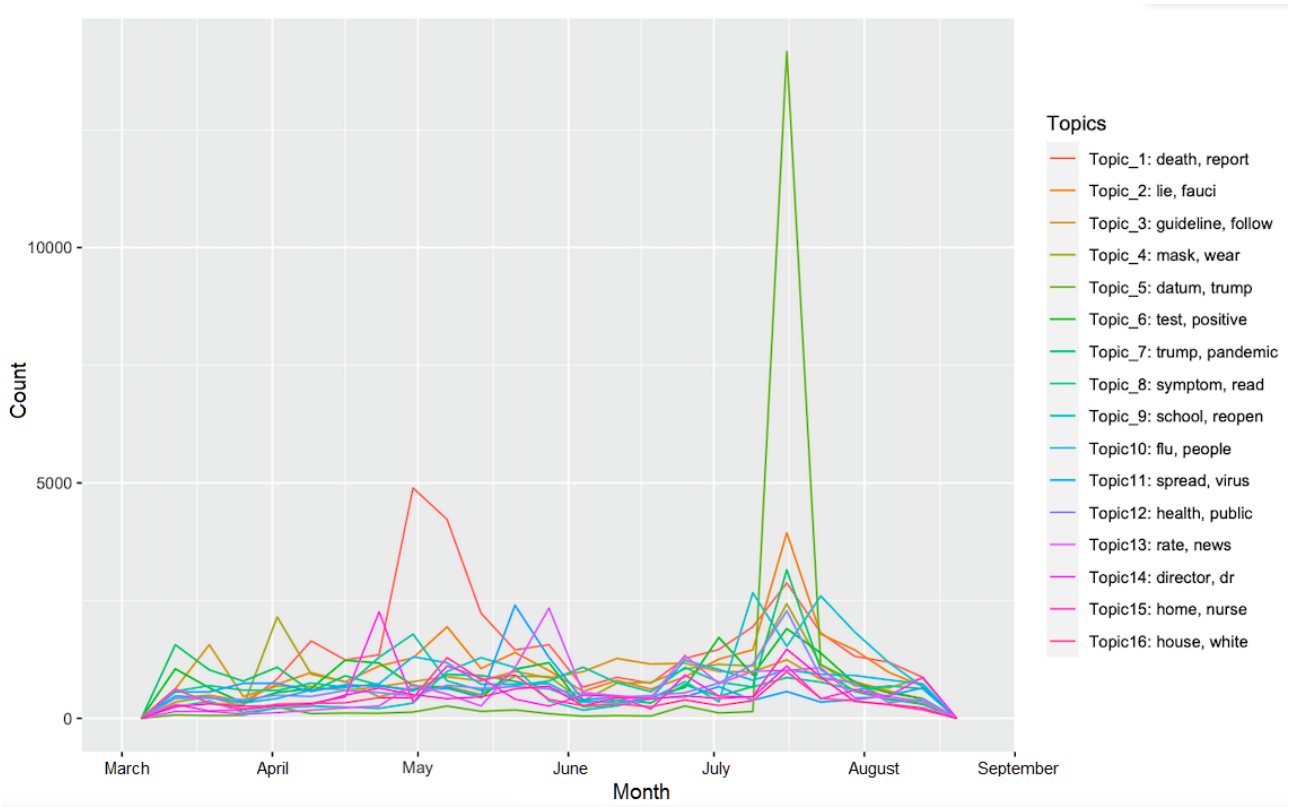


Topic Modeling

Topic modeling provides an automatic, or unsupervised, way of summarizing a large collection of documents. It can help discover hidden themes in the collection, group documents into the discovered themes, and summarize the documents by topic. Topic modeling is often referred to as *soft* clustering, but it is more robust and provides better and more realistic results than typical clustering (eg, k-means clustering) or *hard* clustering [43]. A typical clustering algorithm assumes a distance measure between topics and assigns one topic to each document, whereas topic modeling assigns a document to a collection of topics with different weights or probabilities without any assumption on the distance measure between topics. There are many topic models available. "The most widely used model for topic modeling is the latent Dirichlet allocation (LDA) model" [43], developed by David Blei, Andrew Ng, and Michael I Jordan in 2002 [44].

To extract common topics from this sheer number of tweets, we used the LDA algorithm for topic modeling. We performed the LDA algorithm on the data using the R *textmineR* package, version 3.0.4. The LDA algorithm requires manually inputting the number of expected topics. We ran the LDA algorithm on the data by varying the topic number from 2 through 40. For each topic number, we calculated the coherence score using the *textmineR* package; we ended up choosing 16 topics for the final model, as the topic number that was equal to 16 yielded the highest coherence score. The top eight terms from each of the 16 topics were generated. We also used the *geo_freqpoly* function in the R package *ggplot2*, version 3.3.2, to generate the frequency polygons (see Figure 2) in order to visualize the weekly frequency of the 16 topics from March 11 to August 14, 2020. For each tweet, the LDA assigned a probability to each of the 16 topics. We assigned the topic with the highest probability to a tweet and we grouped the tweets according to the most prevalent topics. To obtain representative tweets for each topic, we randomly sampled 100 tweets from each topic; the two authors then independently examined the sampled tweets, followed by a group discussion to select the most representative ones. If one of the authors thought that there were no conspicuous topics that emerged from the first 100 sampled tweets, another 100 tweets would be sampled and further reviewed; the authors continued this process until the two judged that there was a clear common topic and they reached a consensus. We used the *textmineR* package's topic label function to generate an initial labeling for the topics. After carefully reading through the sampled tweets from each topic, the two authors refined the machine-generated labeling to give each topic the most accurate, concise, and coherent description (see Table 1). Through discussions, the authors further grouped the topics into overarching themes.

Figure 2. Weekly frequency of each topic on Twitter, from March 11 to August 14, 2020.



Results

Overview

The Twitter data generated 16 topics that the public linked to the CDC when they talked about COVID-19. Table 1 shows the topics generated, the number and percentage of each topic, description of topics, and the representative tweets. The topics are sorted according to the number of the tweets in decreasing order. Among the topics, the most discussed was COVID-19 death counts, accounting for 12.16% of the total tweets included in the analysis, followed by general opinions about the credibility of the CDC and other authorities and the CDC's COVID-19 guidelines, with over 20,000 tweets for each topic. The topics in Table 1 can be categorized into four overarching themes, as discussed in the four sections following the table.

Table 1. COVID-19-related tweets about the Centers for Disease Control and Prevention (CDC) by topic.

Topic No.	Terms contributing to topic model	Total tweets (N=290,764), n (%)	Description	Examples of representative tweets (date posted)
1	Death, report, count, week, numb, total, datum, and pneumonia	35,347 (12.16)	Discussion of COVID-19 death counting, focusing on whether the accounting of COVID-19 cases is accurate	"Actually, they aren't even lying about it. The CDC was frank about changing their accounting of COVID-19 cases to encompass contacts with presumptive cases, statistically causing this sudden huge case explosion. [link]" (July 13, 2020) "COVID death counts are inflated. According to CDC: 'ideally testing for COVID-19 should be conducted but it is acceptable to report COVID-19 on a death certificate without this confirmation if the circumstances are compelling within a reasonable degree of certainty.' Read and rt! [link]" (August 7, 2020)
2	Lie, Fauci, people, Trump, trust, doctor, listen, and vaccine	26,026 (8.95)	General opinions about credibility of the CDC and other authorities	"Coronavirus is not political. the curve has not flattened and people are still dying. Trump and our leaders need to listen to scientists and the CDC. You are, most definitely, a cultist." (July 10, 2020)(July 28, 2020) "People: I don't believe anything about the coronavirus I hear from the government, Dr. Fauci, CDC, FDA, WHO, JLA, OPP, or anyone else I've ever heard of. Nope." (July 28, 2020)"
3	Guideline, follow, follow_guideline, people, stay, recommendation, social, and distance	20,032 (6.89)	CDC's COVID-19 guidelines, with a considerable number of tweets about the social distancing and stay-at-home orders	"Bottom line: the more adherent we are to the CDC guidelines, the faster the economy will recover. Stop gathering, put masks on and we will get out of this. Labor secretary: discipline now in following COVID-19 guidelines will end economic slump 'quickly'. [link]" (April 4, 2020) "Coming back to work after 3 months stay-home order (COVID-19) following all safety measures CDC guidelines. Safety and health of staff and clients are utmost importance. #COVID19 #safetyfirst #von #mianailspa. [link]" (June 24, 2020)
4	Mask, wear, wear_mask, spread, recommend, people, cloth, and public	19,934 (6.86)	CDC's recommendation of wearing face masks to slow the spread of virus	"@[tag] @[tag] @[tag] well, here's the us' CDC giving guidelines on how to convert a bandana or even a t-shirt to a face covering. gl! [link]" (April 11, 2020) "CDC recommends people wear cloth face coverings in public settings when around people outside of their household, especially when other social distancing measures are difficult to maintain. [link]" (June 29, 2020)
5	Datum, Trump, hospital, administration send, report, Trump_administration, and control	19,098 (6.57)	Commenting on COVID-19 data reporting being routed away from the CDC	"@[tag] let's make sure coronavirus reporting isn't politicized. If data isn't going to the CDC it needs to be transparent. What a crazy move in the middle of a pandemic!" (July 15, 2020) "This is a very bad thing: white house strips CDC of data collection role for COVID-19 hospitalizations. [link]" (July 15, 2020)
6	Test, positive, antibody, test_positive, result, virus, kit, and antibody_test	18,937 (6.51)	Tweets focusing on the accuracy and inaccuracy of antibody tests	"@[tag] @[tag] CDC? the agency that used COVID19 tainted tests in the beginning of February? Where did CDC get the tests?" (April 22, 2020) "@[tag] @[tag] @[tag] haha from the CDC.... the CDC also say that a positive antibody test may not be 'COVID19' and may be an antibody picked up from a virus such as the common cold.... most of these tests are faulty...." (July 4, 2020)
7	Trump, pandemic, response, fund, test, cut, president, and administration	18,888 (6.50)	Commenting on the Trump administration's health care policies, focusing on cutting CDC funding and dismantling the pandemic response team	"@[tag] @[tag] President Obama created the best pandemic unit in the world. The whole world looked at it as shining example. That pandemic teams' purpose was to secure preparedness for the USA! Republicans fired them in 2018, cut funds for CDC and gave tax breaks to rich!" (March 16, 2020) "The Trump administration is trying to block funding for coronavirus testing and contact tracing, as well as for the CDC, in the upcoming coronavirus relief bill [link]" (July 18, 2020)
8	Symptom, read, update, pandemic, virus, outbreak, guidance, and list	18,326 (6.30)	CDC adding new symptoms of COVID-19 to its list: six symptoms in April and three in June 2020	"The us centers for disease control and prevention has added six new symptoms of COVID-19 to its list: chills, repeated shaking, muscle pain, headache, sore throat, new loss of taste or smell." (April 28, 2020) "New post!!! Follow the link provided World_News CDC: here are 3 'new' COVID-19 coronavirus symptoms to make 12- Forbes [link]" (June 27, 2020)

Topic No.	Terms contributing to topic model	Total tweets (N=290,764), n (%)	Description	Examples of representative tweets (date posted)
9	School, reopen, child, risk, guideline, kid, guidance, and report	18,201 (6.26)	Worries about re-opening schools, and discussion of children's risk of COVID-19	"Exactly why we shouldn't be opening schools before they can meet the CDC requirements. [link]" (July 18, 2020) "And this is what returning to school will look like...260 at Georgia overnight camp test positive for coronavirus, CDC says [link]" (August 2, 2020)
10	Flu, people, die, death, million, American, vaccine, and month	16,950 (5.83)	Comparing the number of COVID-19 deaths with the number of flu deaths	"COVID-19 has led to more than 454,000 illnesses and more than 20,550 deaths worldwide. In the US alone, the flu (also called influenza) has caused an estimated 38 million illnesses, 390,000 hospitalizations and 23,000 deaths this season, according to the CDC." (April 11, 2020) "@[tag] according to the CDC there were between 12k and 61k flu deaths per year over the past 10 years. Taking the worst year (which was an outlier) COVID-19 has killed roughly 3x as many victims in half the time. COVID-19 is not the flu!!" (August 10, 2020)
11	Spread, virus, China, surface, travel, easily, Wuhan, and January	15,223 (5.24)	A wide range of discussion surrounding the spread of COVID-19: China's warning of human-to-human transmission in January 2020, travel restrictions, and surface spread	"Coronavirus 'does not spread easily' by touching surfaces or objects, CDC says. But it still 'may be possible.' [link] via @[tag] possible but not likely - time to get over it. like any germ." (May 21, 2020) "Once people start traveling again, the risk of transmission will surge. 'It keeps me up at night,' CDC's Dr. Cochi in @[tag] about growing immunity gaps for measles, polio and other vaccine-preventable diseases as countries pause vaccination campaigns to mitigate #COVID19 spread [link]" (June 16, 2020)
12	Health, public, datum, public_health, government, report, and agency	15,010 (5.16)	Tweets discussing the CDC not leading the control over COVID-19	"With the white house now having all COVID-19 data and not allowing the CDC to monitor it, the information will no longer be provided to the public. This is a scary time where the government will no longer tell us about the pandemic." (July 16, 2020) "US government health advisers say hospitals are 'scrambling' after Trump administration's 'abrupt' change to COVID-19 data reporting requirements – 'it's another example of CDC being sidelined'. @[tag], told @[tag] [link]" (August 14, 2020)
13	Rate, news, death_rate, death, estimate, low, infection, and report	13,963 (4.80)	Death rate of COVID-19, with a considerable number of tweets emphasizing the low death rate	"Best estimate is 0.4 percent death rate for COVID-19 patients with symptoms: CDC [link]" (May 22, 2020) "The CDC just confirmed that #COVID19 has a 0.2% fatality rate, which is lower than the seasonal flu. Some other news you may have missed while being intentionally distracted [link]" (June 11, 2020)
14	Director, Dr, Redfield, warn, Fauci, wave, bad, and Robert	13,393 (4.61)	Quoting CDC director Dr Robert Redfield's statements, especially the warning of a deadlier second wave	"CDC director warns that COVID-19 could return in winter combined with flu in deadlier second wave - Q13 Fox News [link]" (April 22, 2020) "CDC director: 'The fall and the winter of 2020 and 2021 are going to be the probably one of the most difficult times that we experienced in American public health.'[link]" (July 15, 2020)
15	Home, nurse, patient, health, care, nurse_home, hospital, and official	10,906 (3.75)	Comments on the practice of sending COVID-19 patients to nursing homes	"#reopennj #openamericanow #trump2020 Murphy administration ignored advice and sent COVID-19 patients to nursing homes | mulshine [link]" (May 22, 2020) "5 governors ordered nursing homes to take COVID-19 patients that caused thousands of deaths, for which they now blame CDC and Trump: CA gov. Gavin Newsom. NY gov. Andrew Cuomo. NJ gov. Phil Murphy. MI gov. Gretchen Whitmer. PA gov. Tom Wolf." (June 23, 2020)

Topic No.	Terms contributing to topic model	Total tweets (N=290,764), n (%)	Description	Examples of representative tweets (date posted)
16	House, white, White_House, force, task, task_force, Trump, and CNN	10,530 (3.62)	Tension between the White House and the CDC, focusing on CDC being sidelined in COVID-19 fight	"The White House's coronavirus task force response coordinator, Deborah Birx, said in a recent meeting that 'there is nothing from the CDC that I can trust,' the Washington Post reported. Surprised?" (May 12, 2020) "#trumpviruscoverup #trumpfailedamerica #trumpisalaughingstock Dr. Rich Besser: CDC 'sidelined' from role as leader in #COVID19 fight [link]" (July 18, 2020)

Theme 1: Knowing the Virus and the Situation

Based on the number of tweets, the most tweeted theme was knowing the virus and situation (132,139/290,764, 45.45%). This theme consisted of seven topics: discussion of COVID-19 death counting (35,347/290,764, 12.16%), accuracy of antibody test (18,937/290,764, 6.51%), new symptoms added to the list of COVID-19 (18,326/290,764, 6.30%), number of COVID-19 deaths (16,950/290,764, 5.83%), spread of COVID-19 (15,223/290,764, 5.24%), death rate of COVID-19 (13,963/290,764, 4.80%), and the CDC director Dr Robert Redfield's warning of a deadlier second wave (13,393/290,764, 4.61%). This theme reflected the public's desire to know the virus, such as how it spreads, symptoms of infection, the risk of death, and the situation, such as whether the current response is accurate and effective and how the situation will change. COVID-19 death–related discussion, including death counting, death number, and death rate, dominated this theme.

Theme 2: Policy and Government Actions

The second most tweeted theme was policy and government actions (92,633/290,764, 31.86%). This theme consisted of six topics: commenting on COVID-19 data reporting being routed away from the CDC (19,098/290,764, 6.57%), the Trump administration's health care policies (18,888/290,764, 6.50%), the policy of reopening schools (18,201/290,764, 6.26%), the CDC not leading the control over COVID-19 (15,010/290,764, 5.16%), the practice of sending COVID-19 patients to nursing homes (10,906/290,764, 3.75%), and the tension between the White House and the CDC (10,530/290,764, 3.62%). Tweets under this theme featured comments that challenged the government's actions and policies. Many tweets mentioned that the dismissal of the pandemic response team in 2018 and cutting the CDC's funding weakened the CDC during the COVID-19 pandemic. When the government announced on July 15, 2020, that COVID-19 hospital data would not be reported to the CDC, the number of tweets related to the topic of the CDC not leading the control over COVID-19 for a single day and a single week both set the record (5954 tweets on July 15, 2020, and 13,392 tweets in the week starting on July 15, 2020; see Figure 2 for reference). The dominant voices were complaints against this policy change.

Theme 3: Response Guidelines

The third most tweeted theme was response guidelines (39,966/290,764, 13.75%), which was about how to respond to COVID-19. This theme covered two topics: the CDC's COVID-19 guidelines with a focus on social distancing and stay-at-home orders (20,032/290,764, 6.89%) and the CDC's recommendation of wearing face masks (19,934/290,764, 6.86%). Both of these topics were highly discussed on Twitter, ranking third and fourth, respectively, according to the number of tweets for a single topic. Most of the tweets under this theme suggested that the CDC's guidelines for individuals, businesses, and other organizations should be followed. Many tweets provided the CDC links to the public for further details; one of the most common CDC links mentioned by Twitter users was the video tutorial released by the CDC about making cloth masks.

Theme 4: General Opinion About Credibility

The topic of general opinion about credibility of the CDC and other authorities in charge, such as Dr Fauci, Dr Birx, President Donald Trump, the Food and Drug Administration, and the WHO, stands alone as a category of themes, being the fourth most tweeted theme (26,026/290,764, 8.95%) and the second most tweeted topic, trailing only behind the topic of COVID-19 death counting. Different from the other topics, the tweets under this topic did not point to one or a few specific things; instead they usually expressed general opinions and sometimes together with emotions. Words reflecting "credibility," such as "lie," "trust," "listen," "hoax," "conspiracy," "stupidity," and "fail," were frequently used by Twitter users. However, it was noted that the negative words did not always point to the CDC; instead, there were a substantial number of tweets grouped under this same theme asking people to stand with the CDC and listen to the scientists (see the representative tweets of this topic in Table 1).

Discussion

Principal Findings

Revealed by the quantity of tweets, the public's most prominent concern was death, with over 22.79% of tweets relating to death-related discussion. Previous infoveillance studies of Twitter data in the early period of COVID-19 found that 4.34% of tweets were about death reporting [45] and 10.54% of tweets pertained to deaths caused by COVID-19 [29]. The substantial increase in the death-related discussion with the progression of the pandemic highlighted the urgency of communicating adjusting information to the public, which refers to the information helping them to cope psychologically in threatening situations [46]. Fear and stress were common emotions during the COVID-19 pandemic [29]. Much fear derives from uncertainty and the unknown. Furthermore, the perceived threat in challenging situations motivated people to actively seek information to ease the uncertainty caused by the crisis [47,48]. This explained why a considerable amount of discussion focused

on understanding the COVID-19 virus and how the virus has been coped with. In order to put the impact of COVID-19 into perspective, many tweets compared the death rate of COVID-19 with influenza, H1N1 swine flu, Ebola, and pneumonia, which are more familiar to the public. Discussion about the accuracy of COVID-19 death counting and antibody tests also shows the public's concerns about the current actions of the agencies in charge in response to COVID-19. These findings indicate that in large-scale public health crises such as the COVID-19 pandemic, an imperative component of communication to the public should be informing them of the knowledge of the virus and the factual information about the situations to alleviate fears and confusion. More direct interaction with the public on social media, such as holding an online chat as the CDC had done in the Ebola [20] and Zika outbreaks [13], may also help provide the public with reassurance. In addition, comprehensibility is an important consideration for COVID-19 communication to the public: using language that fits the level of public knowledge helps address the possible misunderstanding of information and avoid the dissemination of misinformation and even rumors.

The majority of the public discussion involved how to act during the COVID-19 pandemic. This echoed past crisis research: in risky environments, the first information that should be conveyed to the public is the information that instructs the public on how to protect themselves in the threatened environments [46]. It also showed that taking actions to prevent the virus from spreading, such as wearing masks and observance of social distancing orders, is a constant topic of the public from the prepandemic period to the peripandemic period [29,45]. The CDC's instructions on how to act in the context of the COVID-19 pandemic, such as guidelines for reopening, recommendations on wearing masks, and how to make masks, have successfully attracted the public attention as soon as they were released. The public not only spread the guidelines widely on Twitter, but they also tweeted explicitly to urge people to follow the CDC's guidelines by providing official CDC links in their tweets. This reflected the public's urgent need for such information to guide their actions, and they took the instructing information from the CDC very seriously. During the unprecedented crisis of COVID-19, scientific understanding of the virus takes time and keeps evolving. Our study suggested that in the next round of COVID-19 communication, the CDC should continue to strive to translate scientific findings into practical instructions, to provide guidance on how to act for both individuals and organizations, and, finally, to protect the public during the pandemic.

As to the CDC's performance in the COVID-19 response, the public expressed mixed comments. One factor contributing to this may be that the CDC has not played a central role in controlling the pandemic; this deviated from what the CDC had done historically during epidemics [49,50]. Even so, it is noted that the discussions on Twitter showed that most of the Twitter users still looked up to the CDC as the authority in disease control and had great expectations for the CDC to lead the fight against COVID-19. While there were negative wordings (eg, "liars," "hoax," "stupidity," and "failed") in the public's general opinions about agencies in charge during the pandemic, including the CDC, it was noticeable that many tweets attributed the current performance of the CDC to the government's policy, criticizing the Trump administration's policies and actions for undermining the functioning of the CDC in response to COVID-19. An early study on the outreach efforts of public health authorities on Facebook found that the spike in public response happens in conjunction with specific events [6]. In our study, the trigger event for a record number of tweets was the announcement that the reporting of COVID-19 hospital data would be sent to the Trump administration rather than the CDC, and the dissenting voice dominated the discussion on this topic. To a large extent, this finding is consistent with the findings of a survey study that showed that Americans' average trust rating for the CDC was significantly higher than that for President Trump [40]. The significance of positive public perception of public health agencies has been receiving increasing recognition [9,10,51,52]. It has been found that greater trust in the CDC was associated with increased knowledge and a lower acceptance of misinformation [40]. The widespread dissemination of the CDC guidelines, as well as the fast speed at which they were circulated, on the one hand reflected the public's urgent need for information as discussed above and, on the other hand, it reflected their trust in the CDC. Even though the CDC's coping so far has not been satisfactory as shown in the tweets, the public's general trust in the CDC is an intangible asset that the CDC can tap into in the next round of the fight against COVID-19.

Limitations

There are a few limitations to this study. First of all, tweets from accounts marked as private might be missed in the data collection, and tweets generated by bots or fake accounts might not have been filtered. Second, this study identified topics from the public discussion about the CDC but did not examine the temporal variance of topics. Although this is not in our research scope, it may deepen our understanding of how the public changed their focus as time and specific situations during that time changed. Therefore, we highly suggest that future studies put emphasis on the temporal dimension of online public discussion about COVID-19 to get more insight into the formation and variation of the discussion topics. Third, this study did not investigate the public's emotions shown in the tweets, which is an important dimension of the public discussion. Future research in this line of study may shed light on the public's affective response to the CDC's actions and may inform the CDC about the public's emotions to be addressed during the pandemic. Lastly, Twitter users do not represent the US population [20]. Therefore, as with all social media analyses, findings of this study cannot be generalized to the whole American public.

Conclusions

In public health crises, social media platforms, such as Twitter, can provide valuable databases for public health agencies to understand the public's concerns, focus of attention, and expectations. The ability of text mining to derive high-quality information from massive data sets is ideal for performing surveillance work. Especially in a protracted pandemic such as COVID-19, quickly and efficiently identifying the topics within the public discussion on Twitter would provide insight for the

next round of public health communication in order to mitigate public concerns and avoid the spread of misinformation.

Acknowledgments

This study was supported by the National Cancer Institute Grant T32 CA 113710.

Conflicts of Interest

None declared.

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2020-08-11]
2. Countries where COVID-19 has spread. Worldometer. 2020. URL: https://www.worldometers.info/coronavirus/countries-where-coronavirus-has-spread/ [accessed 2020-08-05]
3. Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First case of 2019 novel coronavirus in the United States. N Engl J Med 2020 Mar 05;382(10):929-936. [doi: 10.1056/nejmoa2001191]
4. COVID Data Tracker. Centers for Disease Control and Prevention. 2020. URL: https://covid.cdc.gov/covid-data-tracker/ [accessed 2020-08-05]
5. CDC. About CDC 24-7. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/about/default.htm [accessed 2020-08-05]
6. Sesagiri Raamkumar A, Tan SG, Wee HL. Measuring the outreach efforts of public health authorities and the public response on Facebook during the COVID-19 pandemic in early 2020: Cross-country comparison. J Med Internet Res 2020 May 19;22(5):e19334 [FREE Full text] [doi: 10.2196/19334] [Medline: 32401219]
7. Gilson L. Trust and the development of health care as a social institution. Soc Sci Med 2003 Apr;56(7):1453-1468. [doi: 10.1016/s0277-9536(02)00142-9]
8. Larson HJ, Heymann DL. Public health response to influenza A(H1N1) as an opportunity to build public trust. JAMA 2010 Jan 20;303(3):271-272. [doi: 10.1001/jama.2009.2023] [Medline: 20085957]
9. Wise K. Why public health needs relationship management. J Health Hum Serv Adm 2008;31(3):309-331. [Medline: 19209562]
10. Springston J, Weaver Lariscy RA. Public relations effectiveness in public health institutions. J Health Hum Serv Adm 2005;28(2):218-245. [Medline: 16521670]
11. Pitrelli N, Sturloni G. Infectious diseases and governance of global risks through public communication and participation. Ann Ist Super Sanita 2007;43(4):336-343. [Medline: 18209267]
12. Wynia MK. Risk and trust in public health: A cautionary tale. Am J Bioeth 2006;6(2):3-6. [doi: 10.1080/15265160500508635] [Medline: 16500837]
13. Glowacki EM, Lazard AJ, Wilcox GB, Mackert M, Bernhardt JM. Identifying the public's concerns and the Centers for Disease Control and Prevention's reactions during a health crisis: An analysis of a Zika live Twitter chat. Am J Infect Control 2016 Dec 01;44(12):1709-1711. [doi: 10.1016/j.ajic.2016.05.025] [Medline: 27544795]
14. Joob B, Wiwanitkit V. Zika live Twitter chat. Am J Infect Control 2016 Dec 01;44(12):1756-1757. [doi: 10.1016/j.ajic.2016.08.019] [Medline: 27751615]
15. Miller M, Banerjee T, Muppalla R, Romine W, Sheth A. What are people tweeting about Zika? An exploratory study concerning its symptoms, treatment, transmission, and prevention. JMIR Public Health Surveill 2017 Jul 19;3(2):e38 [FREE Full text] [doi: 10.2196/publichealth.7157] [Medline: 28630032]
16. Stefanidis A, Vraga E, Lamprianidis G, Radzikowski J, Delamater PL, Jacobsen KH, et al. Zika in Twitter: Temporal variations of locations, actors, and concepts. JMIR Public Health Surveill 2017 May 20;3(2):e22 [FREE Full text] [doi: 10.2196/publichealth.6925] [Medline: 28428164]
17. Chen S, Xu Q, Buchenberger J, Bagavathi A, Fair G, Shaikh S, et al. Dynamics of health agency response and public engagement in public health emergency: A case study of CDC tweeting patterns during the 2016 Zika epidemic. JMIR Public Health Surveill 2018 Dec 22;4(4):e10827 [FREE Full text] [doi: 10.2196/10827] [Medline: 30467106]
18. Allem J, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-cigarette surveillance with social media data: Social bots, emerging topics, and trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]
19. Glowacki EM, Glowacki JB, Chung AD, Wilcox GB. Reactions to foodborne Escherichia coli outbreaks: A text-mining analysis of the public's response. Am J Infect Control 2019 Oct;47(10):1280-1282. [doi: 10.1016/j.ajic.2019.04.004] [Medline: 31104869]

20. Lazard AJ, Scheinfeld E, Bernhardt JM, Wilcox GB, Suran M. Detecting themes of public concern: A text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. Am J Infect Control 2015 Oct 01;43(10):1109-1111. [doi: 10.1016/j.ajic.2015.05.025] [Medline: 26138998]
21. Mamidi R, Miller M, Banerjee T, Romine W, Sheth A. Identifying key topics bearing negative sentiment on Twitter: Insights concerning the 2015-2016 Zika epidemic. JMIR Public Health Surveill 2019 Jul 04;5(2):e11036 [FREE Full text] [doi: 10.2196/11036] [Medline: 31165711]
22. de Araujo DHM, de Carvalho EA, da Motta CLR, da Silva Borges MR, Gomes JO, de Carvalho PVR. Social networks applied to Zika and H1N1 epidemics: A systematic review. In: Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018). Cham, Switzerland: Springer; 2018 Presented at: 20th Congress of the International Ergonomics Association (IEA 2018); August 26-30, 2018; Florence, Italy p. 679-692. [doi: 10.1007/978-3-319-96089-0_74]
23. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic-Analysis of Facebook as a social media health information platform. Am J Infect Control 2017 Mar 01;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
24. Paul MJ, Dredze M, Broniatowski D. Twitter improves influenza forecasting. PLoS Curr 2014 Oct 28;6 [FREE Full text] [doi: 10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117] [Medline: 25642377]
25. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Comput Biol 2015 Oct;11(10):e1004513 [FREE Full text] [doi: 10.1371/journal.pcbi.1004513] [Medline: 26513245]
26. Harris JK, Hawkins JB, Nguyen L, Nsoesie EO, Tuli G, Mansour R, et al. Using Twitter to identify and respond to food poisoning: The food safety STL project. J Public Health Manag Pract 2017;23(6):577-580 [FREE Full text] [doi: 10.1097/PHH.0000000000000516] [Medline: 28166175]
27. Guo J, Radloff CL, Wawrzynski SE, Cloyes KG. Mining Twitter to explore the emergence of COVID-19 symptoms. Public Health Nurs 2020 Nov;37(6):934-940. [doi: 10.1111/phn.12809] [Medline: 32937679]
28. Social media at CDC: Twitter. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/socialmedia/tools/Twitter.html [accessed 2020-08-05]
29. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: Infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
30. Li J, Xu Q, Cuomo R, Purushothaman V, Mackey T. Data mining and content analysis of the Chinese social media platform Weibo during the early COVID-19 outbreak: Retrospective observational infoveillance study. JMIR Public Health Surveill 2020 Apr 21;6(2):e18700 [FREE Full text] [doi: 10.2196/18700] [Medline: 32293582]
31. Shen C, Chen A, Luo C, Zhang J, Feng B, Liao W. Using reports of symptoms and diagnoses on social media to predict COVID-19 case counts in mainland China: Observational infoveillance study. J Med Internet Res 2020 May 28;22(5):e19421 [FREE Full text] [doi: 10.2196/19421] [Medline: 32452804]
32. Sarker A, Lakamana S, Hogg-Bremer W, Xie A, Al-Garadi M, Yang Y. Self-reported COVID-19 symptoms on Twitter: An analysis and a research resource. J Am Med Inform Assoc 2020 Aug 01;27(8):1310-1315 [FREE Full text] [doi: 10.1093/jamia/ocaa116] [Medline: 32620975]
33. Tao Z, Chu G, McGrath C, Hua F, Leung YY, Yang W, et al. Nature and diffusion of COVID-19-related oral health information on Chinese social media: Analysis of tweets on Weibo. J Med Internet Res 2020 Jun 15;22(6):e19981 [FREE Full text] [doi: 10.2196/19981] [Medline: 32501808]
34. Wahbeh A, Nasralah T, Al-Ramahi M, El-Gayar O. Mining physicians' opinions on social media to obtain insights into COVID-19: Mixed methods analysis. JMIR Public Health Surveill 2020 Jun 18;6(2):e19276 [FREE Full text] [doi: 10.2196/19276] [Medline: 32421686]
35. Budhwani H, Sun R. Creating COVID-19 stigma by referencing the novel coronavirus as the "Chinese virus" on Twitter: Quantitative analysis of social media data. J Med Internet Res 2020 May 06;22(5):e19301 [FREE Full text] [doi: 10.2196/19301] [Medline: 32343669]
36. Rufai S, Bunce C. World leaders' usage of Twitter in response to the COVID-19 pandemic: A content analysis. J Public Health (Oxf) 2020 Aug 18;42(3):510-516 [FREE Full text] [doi: 10.1093/pubmed/fdaa049] [Medline: 32309854]
37. Park HW, Park S, Chong M. Conversations and medical news frames on Twitter: Infodemiological study on COVID-19 in South Korea. J Med Internet Res 2020 May 05;22(5):e18897 [FREE Full text] [doi: 10.2196/18897] [Medline: 32325426]
38. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
39. Pobiruchin M, Zowalla R, Wiesner M. Temporal and location variations, and link categories for the dissemination of COVID-19-related information on Twitter during the SARS-CoV-2 outbreak in Europe: Infoveillance study. J Med Internet Res 2020 Aug 28;22(8):e19629 [FREE Full text] [doi: 10.2196/19629] [Medline: 32790641]
40. Dhanani LY, Franz B. The role of news consumption and trust in public health leadership in shaping COVID-19 knowledge and prejudice. Front Psychol 2020;11:560828 [FREE Full text] [doi: 10.3389/fpsyg.2020.560828] [Medline: 33192827]

41. Banda JM, Tekumalla R, Wang G, Yu J, Liu T, Ding Y, et al. A large-scale COVID-19 Twitter chatter dataset for open scientific research - An international collaboration (Version 32) [Data set]. Zenodo. 2020 Oct 18. URL: http://doi.org/10.5281/zenodo.4104587 [accessed 2021-01-31]
42. The Panacea Lab: COVID-19 Twitter. GitHub. 2020. URL: https://github.com/thepanacealab/covid19_twitter [accessed 2021-02-02]
43. Blum A, Hopcroft J, Kannan R. Foundations of Data Science. Cambridge, UK: Cambridge University Press; 2020.
44. Blei DM, Ng A, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003 Mar;3:993-1022 [FREE Full text] [doi: 10.1162/jmlr.2003.3.4-5.993]
45. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study. J Med Internet Res 2020 Oct 23;22(10):e22624 [FREE Full text] [doi: 10.2196/22624] [Medline: 33006937]
46. Sturges DL. Communicating through crisis. Manag Commun Q 2016 Aug 15;7(3):297-316. [doi: 10.1177/0893318994007003004]
47. Loges WE. Canaries in the coal mine: Perceptions of threat and media system dependency relations. Communic Res 1994;21(1):5-23. [doi: 10.1177/009365094021001002]
48. Lyu JC. How young Chinese depend on the media during public health crises? A comparative perspective. Public Relat Rev 2012 Dec;38(5):799-806. [doi: 10.1016/j.pubrev.2012.07.006]
49. Greenfield-Boyce N. As the coronavirus crisis heats up, why isn't America hearing from the CDC? NPR. 2020 Mar 25. URL: https://www.npr.org/sections/health-shots/2020/03/25/821009072/ [accessed 2020-11-21]
50. Sun LH. CDC, the top US public health agency, is sidelined during coronavirus pandemic. The Washington Post. 2020 Mar 19. URL: https://www.washingtonpost.com/health/2020/03/19/cdc-top-us-public-health-agency-is-sidelined-during-coronavirus-pandemic/ [accessed 2020-11-21]
51. Glik DC. Risk communication for public health emergencies. Annu Rev Public Health 2007;28:33-54. [doi: 10.1146/annurev.publhealth.28.021406.144123] [Medline: 17222081]
52. Holmes BJ. Communicating about emerging infectious disease: The importance of research. Health Risk Soc 2008 Aug;10(4):349-360. [doi: 10.1080/13698570802166431]

Abbreviations

ASCII: American Standard Code for Information Interchange
CDC: Centers for Disease Control and Prevention
GUI: graphical user interface
JSON: JavaScript Object Notation
LDA: latent Dirichlet allocation
WHO: World Health Organization

Edited by G Eysenbach; submitted 25.10.20; peer-reviewed by A Sesagiri Raamkumar, R Zowalla, A Selya; comments to author 17.11.20; revised version received 24.11.20; accepted 25.01.21; published 09.02.21.

Please cite as:
Lyu JC, Luli GK
Understanding the Public Discussion About the Centers for Disease Control and Prevention During the COVID-19 Pandemic Using Twitter Data: Text Mining Analysis Study
J Med Internet Res 2021;23(2):e25108
URL: http://www.jmir.org/2021/2/e25108/
doi:10.2196/25108
PMID:33497351

Review

Efficiency and Quality of Data Collection Among Public Mental Health Surveys Conducted During the COVID-19 Pandemic: Systematic Review

Yu-Hsuan Lin[1,2,3,4*], MD, PhD; Chung-Yen Chen[5,6*], MD; Shiow-Ing Wu[1], DDS, MPH, PhD

[1]Institute of Population Health Sciences, National Health Research Institutes, Miaoli County, Taiwan

[2]Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan

[3]Department of Psychiatry, College of Medicine, National Taiwan University, Taipei, Taiwan

[4]Institute of Health Behaviors and Community Sciences, College of Public Health, National Taiwan University, Taipei, Taiwan

[5]Department of Environmental and Occupational Medicine, National Taiwan University Hospital, Taipei, Taiwan

[6]Institute of Environmental and Occupational Health Sciences, College of Public Health, National Taiwan University, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Yu-Hsuan Lin, MD, PhD
Institute of Population Health Sciences
National Health Research Institutes
35 Keyan Road, Zhunan
Miaoli County, 35053
Taiwan
Phone: 886 37 206166 ext 36383
Email: yuhsuanlin@nhri.edu.tw

Abstract

Background: The World Health Organization has recognized the importance of assessing population-level mental health during the COVID-19 pandemic. During a global crisis such as the COVID-19 pandemic, a timely surveillance method is urgently needed to track the impact on public mental health.

Objective: This brief systematic review focused on the efficiency and quality of data collection of studies conducted during the COVID-19 pandemic.

Methods: We searched the PubMed database using the following search strings: ((COVID-19) OR (SARS-CoV-2)) AND ((Mental health) OR (psychological) OR (psychiatry)). We screened the titles, abstracts, and texts of the published papers to exclude irrelevant studies. We used the Newcastle-Ottawa Scale to evaluate the quality of each research paper.

Results: Our search yielded 37 relevant mental health surveys of the general public that were conducted during the COVID-19 pandemic, as of July 10, 2020. All these public mental health surveys were cross-sectional in design, and the journals efficiently made these articles available online in an average of 18.7 (range 1-64) days from the date they were received. The average duration of recruitment periods was 9.2 (range 2-35) days, and the average sample size was 5137 (range 100-56,679). However, 73% (27/37) of the selected studies had Newcastle-Ottawa Scale scores of <3 points, which suggests that these studies are of very low quality for inclusion in a meta-analysis.

Conclusions: The studies examined in this systematic review used an efficient data collection method, but there was a high risk of bias, in general, among the existing public mental health surveys. Therefore, following recommendations to avoid selection bias, or employing novel methodologies considering both a longitudinal design and high temporal resolution, would help provide a strong basis for the formation of national mental health policies.

(J Med Internet Res 2021;23(2):e25118) doi:10.2196/25118

KEYWORDS

COVID-19; mental health; Newcastle-Ottawa Scale; review; data collection; survey; surveillance; literature; research

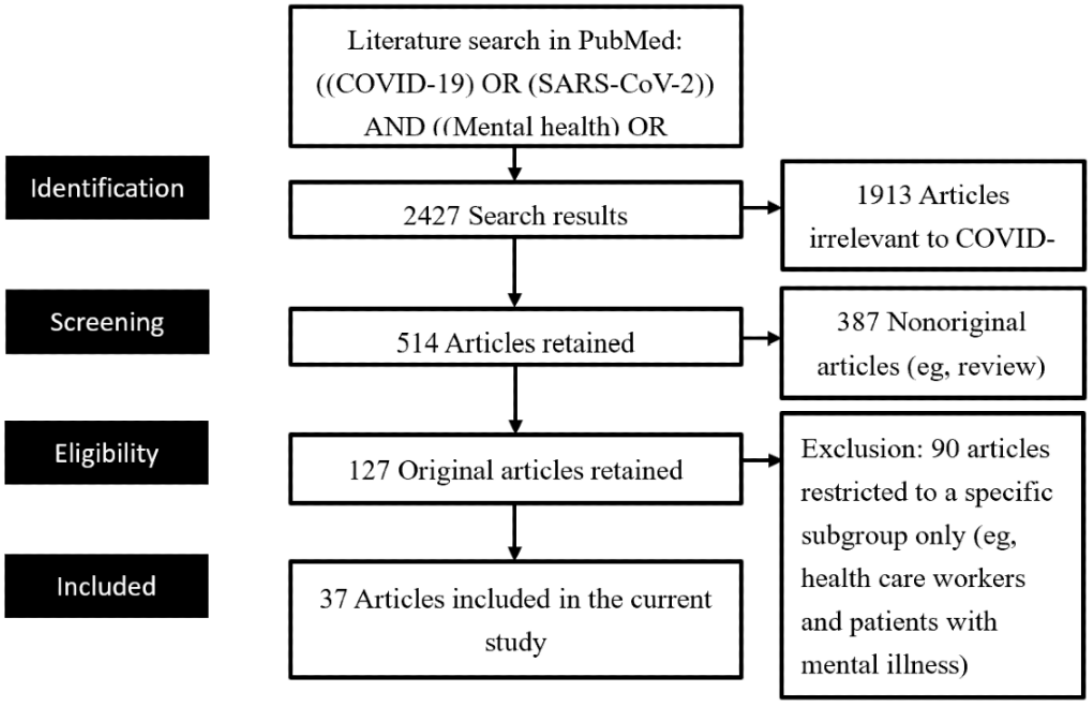
Introduction

The World Health Organization has recognized the importance of assessing population-level mental health during the COVID-19 pandemic. More than 463 articles on the mental health impact of COVID-19 have been published in 2020, and several more mental health surveys are underway [1]. Most of these studies evaluated the psychiatric symptoms among the general public and health care workers, and some others evaluated patients with COVID-19 or those with psychiatric disorders [2]. In a global crisis such as the COVID-19 pandemic, time-sensitive policy decision-making underscores the importance of fostering an agile empirical approach that can monitor population-level mental health in a timely manner. Any change in the mental health impact of COVID-19 is likely to be dynamic. For example, the 20% decrease in suicide rate in Japan observed during the early stage of the pandemic seemed to have reversed in August 2020, when a 7.7% rise was reported [3]. Therefore, a timely surveillance method is urgently needed to track the impact of COVID-19 on public mental health. This brief systematic review focused on studies on the general public conducted during the COVID-19 pandemic and analyzed the efficiency and quality of data collection in the existing literature.

Methods

We used the following search strings to select relevant articles in the PubMed database: ((COVID-19) OR (SARS-CoV-2)) AND ((Mental health) OR (psychological) OR (psychiatry)). We screened the titles, abstracts, and texts of the published articles to exclude irrelevant studies. We also excluded articles published on preprint platforms such as bioRxiv, because we were unable to analyze the time between the receipt of the articles and their availability online. A PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram detailing the study retrieval process is shown in Figure 1.

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



We used the Newcastle-Ottawa Scale to evaluate the quality of each research article. The Newcastle-Ottawa Scale includes four criteria for selection bias, with scores ranging between 0 and 5; scores <3 points are regarded as high risk for selection bias in a cross-sectional study [1,4]. The four quality assessment criteria were as follows: sample representativeness (1 point), predetermined and satisfactory sample size (1 point), comparability between respondents and nonrespondents (1 point), and ascertainment of the measurement tool (2 points for the validated measurement tool; 1 point for the nonvalidated measurement tool, but the tool is available or described).

Results

In all, we found 37 relevant mental health surveys of the general public during the COVID-19 pandemic by searching the PubMed database, as of July 10, 2020. All these public mental health survey studies examined had a cross-sectional study design. The journals efficiently made these articles available online in an average of 18.7 (range 1-64) days from the date they were received. The average duration of recruitment period of these studies was 9.2 (range 2-35) days, and the average sample size was 5137 (range 100-56,679). The correlation coefficient (r=0.164, P=.40) between these durations and the corresponding sample sizes suggested that the samples in the studies selected from the existing literature were collected in an efficient manner; even larger sample sizes did not require proportionately longer to collect the data. Although there was a potential publication bias in the recruitment or data collection period, it is noteworthy that approximately 92% (34/37) of the studies reported the use of a web-based or an app-based survey,

thereby promoting the efficiency of large-scale data collection, especially during the COVID-19 pandemic.

Unlike most studies on health care workers, which have a low risk of selection bias [1], our review showed that 73% (27/37) of the selected studies evaluating the mental health impact of COVID-19 on the general public had Newcastle-Ottawa Scale scores of <3 points, which suggests that these studies were of very low quality to be considered for inclusion in a meta-analysis. All of the selected studies met the criterion of validated measurement tools, but only 8 studies met the criterion of sample representativeness, 5 met the criterion of justified sample size, and 1 met the criterion of comparability between respondents and nonrespondents (Table 1).

Table 1. Public mental health surveys conducted during the COVID-19 pandemic.

Publication	Sample size (n)	Recruitment period (2020)	Duration of recruitment period (days)	Duration between article submission and online publication (days)	Newcastle-Ottawa Quality Assessment Scale scores				
					Sample representativeness	Sample size	Nonrespondents	Ascertainment of exposure	Total score
Ahmed et al [5]	1074	__[a]	—	32	0	0	0	2	2
Bryan et al [6]	10,625	March 18 to April 4	18	—	1	1	0	2	4
Cao et al [7]	7143	—	—	6	1	0	1	2	4
Castelli et al [8]	1321	March 19 to April 5	18	—	0	0	0	2	2
Choi et al [9]	500	April 24 to May 3	10	18	1	1	0	2	4
Cortes-Alvarez et al [10]	1105	March 30 to April 5	7	—	0	0	0	2	2
Fitzpatrick et al [11]	10,368	March 23-29	7	—	1	0	0	2	3
Gao et al [12]	4827	January 31 to February 2	3	43	0	0	0	2	2
Huang and Zhao [13]	7236	February 3-17	15	23	0	0	0	2	2
Lei et al [14]	1593	February 4-10	7	25	0	0	0	2	2
Li et al [15]	4607	February 2-9	8	—	0	0	0	2	2
Liang et al [16]	584	January 30 to unknown	—	—	0	0	0	2	2
Liu et al [17]	285	January 30 to February 8	10	1	0	0	0	2	2
Ma et al [18]	728	April 9-30	22	—	0	0	0	2	2
Mazza et al [19]	2766	March 18-22	5	19	0	0	0	2	2
McGinty et al [20]	1468	April 7-13	7	—	1	0	0	2	3
Moccia et al [21]	500	April 10-13	4	5	1	1	0	2	4
Moghanibashi-Mansourieh [22]	10,754	March 1-9	9	22	0	0	0	2	2
Naser et al [23]	4126	March 22-28	7	64	0	1	0	2	3
Ozdin and Bayrak Özdin [24]	343	April 14-16	3	—	0	1	0	2	3
Qiu et al [25]	52,730	January 31 to February 10	11	7	0	0	0	2	2
Rehman et al [26]	403	April 3-6	4	49	0	0	0	2	2
Roy et al [27]	662	March 22-24	3	7	0	0	0	1	1
Smith et al [28]	932	March 17 to unknown	—	25	0	0	0	2	2
Shi et al [29]	56,679	February 28 to March 11	13	—	0	0	0	2	2
Sonderskov et al [30]	2458	March 31 to April 6	7	13	1	0	0	2	3
Tan et al [31]	673	February 24-25	2	6	0	0	0	2	2
Tian et al [32]	1060	January 31 to February 2	3	15	0	0	0	2	2

Publication	Sample size (n)	Recruitment period (2020)	Duration of recruitment period (days)	Duration between article submission and online publication (days)	Newcastle-Ottawa Quality Assessment Scale scores				
					Sample representativeness	Sample size	Nonrespondents	Ascertainment of exposure	Total score
Ustun [33]	1115	March 23 to April 3	12	—	0	0	0	2	2
Verma and Mishra [34]	354	April 4-14	11	—	0	0	0	2	2
Wang et al [35]									
	1210	January 31 to February 2	3	4	0	0	0	2	2
	861	February 28 to March 1							
Xiao et al [36]	170	—	—	11	0	0	0	2	2
Yuan et al [37]	939	—	—	22	0	0	0	2	2
Yuan et al [38]	100	—	—	—	0	0	0	2	2
Zhang and Ma [39]	263	January 28 to February 5	9	20	0	0	0	2	2
Zhang et al [40]	369	February 20-21	2	10	1	0	0	2	3
Zhu et al [41]	2279	February 12 to March 17	35	1	0	0	0	2	2

[a]Data not available.

Discussion

The general public surveys examined in this review most commonly used the web-based format to promote efficiency of large-scale data collection, especially in the countries with significant COVID-19 outbreaks that had introduced social distancing or lockdown measures. However, the web-based survey approach has an inherent methodological limitation to show the comparability between respondents and nonrespondents. The sample representativeness of the target population is critical for selection bias. Web-based survey platforms using sampling methods could avoid such selection bias. Two published surveys [6,11] used Qualtrics Panels, a web-based survey platform that uses quota sampling methods to identify participants who meet a given study's eligibility criteria, with a target to recruit a sample that is demographically similar to the 2010 United States census distributions for age, sex, and race or ethnicity (with a ±10% variation). Qualtrics Panels maintains a database of several million US residents who have volunteered to participate in periodic survey-based research. Another study [20] used AmeriSpeak, a probability-based research panel designed to be representative of the adult US population. AmeriSpeak data is sourced from National Opinion Research Center's area probability sample and a US postal service address–based sample covering approximately 97% of all US households. In this national survey study [20], a higher proportion (13.6%) of the US adult sample reported symptoms of serious psychological distress in 2020 than in 2018 (3.9%).

The limitations of cross-sectional studies should also be noted when interpreting the role of the COVID-19 outbreak on the mental health of the general public. These epidemiological studies identified general risk factors during the COVID-19 outbreak, rather than the specific impacts of COVID-19 on the mental health of the public. For example, a cross-sectional study of 4827 participants [12] showed that lower education levels and higher social media exposure result in a high risk of individuals developing depressive symptoms, but these findings did not suggest that COVID-19 had an impact on the public's mental health. There was only one longitudinal study [35], which conducted two-wave surveys during the initial COVID-19 outbreak (ie, January 31 to February 2, 2020) and during the peak 4 weeks later (ie, February 28 to March 1). Although this study recruited 1210 and 861 participants within 3 days by using a web-based survey approach, no significant longitudinal changes in stress, anxiety, or depression levels were observed. These findings might result from a lack of baseline data, few follow-ups with the same respondents (333/1210, 27.52%), or insufficient temporal resolution to detect mental health changes. Only one study compared the prevalence of symptoms of serious psychological distress before and during the COVID-19 pandemic by using an identical measure. This survey, which used AmeriSpeak for sample recruitment, showed that the prevalence of serious psychological distress symptoms among US adults was higher during the COVID-19 pandemic in 2020 than that reported in the 2018 National Health Interview Survey [20]. However, this study did not monitor the changes in population-level mental health through the unfolding pandemic. Timely mental health survey studies with a high temporal

resolution, such as internet search data (eg, Google Trends) [42] and sentiment analysis on social media (eg, Weibo posts) [43], are warranted in the future to monitor the long-term impacts of the fast-moving COVID-19 outbreak.

In conclusion, this systematic review found that the data collection methods used in the public mental health surveys in the existing literature were efficient but generally had a high risk of bias. Therefore, following recommendations to avoid selection bias, or to apply novel methodologies considering both longitudinal design and high temporal resolution, would help provide a strong basis for the formation of national mental health policies.

Acknowledgments

The authors thank Hsiao-Han Lin and Ting-Chen Hsu for their excellent technical assistance during this study. This study was supported by grants from National Health Research Institute in Taiwan (10A1-PHPP22-014; 10A1-PHPP16-014).

Authors' Contributions

YHL, CYC, and SIW conceptualized the study, and analyzed and interpreted the data. YHL and CYC drafted the manuscript. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Pappa S, Ntella V, Giannakas T, Giannakoulis VG, Papoutsi E, Katsaounou P. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: A systematic review and meta-analysis. Brain, Behavior, and Immunity 2020 Aug;88:901-907. [doi: 10.1016/j.bbi.2020.05.026]
2. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. Brain, Behavior, and Immunity 2020 Oct;89:531-542. [doi: 10.1016/j.bbi.2020.05.048]
3. Ueda M, Nordström R, Matsubayashi T. Suicide and mental health during the COVID-19 pandemic in Japan. medRxiv. Preprint posted online on December 20, 2020. [doi: 10.1101/2020.10.06.20207530]
4. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol 2010 Jul 22;25(9):603-605. [doi: 10.1007/s10654-010-9491-z] [Medline: 20652370]
5. Ahmed MZ, Ahmed O, Aibao Z, Hanbin S, Siyu L, Ahmad A. Epidemic of COVID-19 in China and associated psychological problems. Asian J Psychiatr 2020 Jun;51:102092 [FREE Full text] [doi: 10.1016/j.ajp.2020.102092] [Medline: 32315963]
6. Bryan CJ, Bryan AO, Baker JC. Associations among state-level physical distancing measures and suicidal thoughts and behaviors among U.S. adults during the early COVID-19 pandemic. Suicide Life Threat Behav 2020 Dec;50(6):1223-1229 [FREE Full text] [doi: 10.1111/sltb.12653] [Medline: 32589801]
7. Cao W, Fang Z, Hou G, Han M, Xu X, Dong J, et al. The psychological impact of the COVID-19 epidemic on college students in China. Psychiatry Res 2020 May;287:112934 [FREE Full text] [doi: 10.1016/j.psychres.2020.112934] [Medline: 32229390]
8. Castelli L, Di Tella M, Benfante A, Romeo A. The spread of COVID-19 in the Italian population: anxiety, depression, and post-traumatic stress symptoms. Can J Psychiatry 2020 Oct;65(10):731-732. [doi: 10.1177/0706743720938598] [Medline: 32588644]
9. Choi EPH, Hui BPH, Wan EYF. Depression and anxiety in Hong Kong during COVID-19. Int J Environ Res Public Health 2020 May 25;17(10):3740 [FREE Full text] [doi: 10.3390/ijerph17103740] [Medline: 32466251]
10. Cortés-Álvarez NY, Piñeiro-Lamas R, Vuelvas-Olmos CR. Psychological effects and associated factors of COVID-19 in a Mexican sample. Disaster Med Public Health Prep 2020 Jun;14(3):413-424 [FREE Full text] [doi: 10.1017/dmp.2020.215] [Medline: 32576317]
11. Fitzpatrick KM, Harris C, Drawve G. How bad is it? Suicidality in the middle of the COVID-19 pandemic. Suicide Life Threat Behav 2020 Dec;50(6):1241-1249 [FREE Full text] [doi: 10.1111/sltb.12655] [Medline: 32589799]
12. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
13. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]
14. Lei L, Huang X, Zhang S, Yang J, Yang L, Xu M. Comparison of prevalence and associated factors of anxiety and depression among people affected by versus people unaffected by quarantine during the COVID-19 epidemic in Southwestern China. Med Sci Monit 2020 Apr 20;26:e924609. [doi: 10.12659/msm.924609]

15. Li G, Miao J, Wang H, Xu S, Sun W, Fan Y, et al. Psychological impact on women health workers involved in COVID-19 outbreak in Wuhan: a cross-sectional study. J Neurol Neurosurg Psychiatry 2020 Aug;91(8):895-897. [doi: 10.1136/jnnp-2020-323134] [Medline: 32366684]
16. Liang L, Ren H, Cao R, Hu Y, Qin Z, Li C, et al. The effect of COVID-19 on youth mental health. Psychiatr Q 2020 Sep;91(3):841-852 [FREE Full text] [doi: 10.1007/s11126-020-09744-3] [Medline: 32319041]
17. Liu N, Zhang F, Wei C, Jia Y, Shang Z, Sun L, et al. Prevalence and predictors of PTSS during COVID-19 outbreak in China hardest-hit areas: Gender differences matter. Psychiatry Res 2020 May;287:112921 [FREE Full text] [doi: 10.1016/j.psychres.2020.112921] [Medline: 32240896]
18. Ma ZF, Zhang Y, Luo X, Li X, Li Y, Liu S, et al. Increased stressful impact among general population in mainland China amid the COVID-19 pandemic: A nationwide cross-sectional study conducted after Wuhan city's travel ban was lifted. Int J Soc Psychiatry 2020 Dec;66(8):770-779 [FREE Full text] [doi: 10.1177/0020764020935489] [Medline: 32564637]
19. Mazza C, Ricci E, Biondi S, Colasanti M, Ferracuti S, Napoli C, et al. A nationwide survey of psychological distress among Italian people during the COVID-19 pandemic: immediate psychological responses and associated factors. Int J Environ Res Public Health 2020 May 02;17(9):3165 [FREE Full text] [doi: 10.3390/ijerph17093165] [Medline: 32370116]
20. McGinty EE, Presskreischer R, Han H, Barry CL. Psychological distress and loneliness reported by US adults in 2018 and April 2020. JAMA 2020 Jul 07;324(1):93. [doi: 10.1001/jama.2020.9740]
21. Moccia L, Janiri D, Pepe M, Dattoli L, Molinaro M, De Martin V, et al. Affective temperament, attachment style, and the psychological impact of the COVID-19 outbreak: an early report on the Italian general population. Brain Behav Immun 2020 Jul;87:75-79 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.048] [Medline: 32325098]
22. Moghanibashi-Mansourieh A. Assessing the anxiety level of Iranian general population during COVID-19 outbreak. Asian J Psychiatr 2020 Jun;51:102076 [FREE Full text] [doi: 10.1016/j.ajp.2020.102076] [Medline: 32334409]
23. Naser AY, Dahmash EZ, Al-Rousan R, Alwafi H, Alrawashdeh HM, Ghoul I, et al. Mental health status of the general population, healthcare professionals, and university students during 2019 coronavirus disease outbreak in Jordan: A cross-sectional study. Brain Behav 2020 Aug;10(8):e01730 [FREE Full text] [doi: 10.1002/brb3.1730] [Medline: 32578943]
24. Özdin S, Bayrak Özdin Ş. Levels and predictors of anxiety, depression and health anxiety during COVID-19 pandemic in Turkish society: The importance of gender. Int J Soc Psychiatry 2020 Aug;66(5):504-511 [FREE Full text] [doi: 10.1177/0020764020927051] [Medline: 32380879]
25. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. Gen Psychiatr 2020;33(2):e100213 [FREE Full text] [doi: 10.1136/gpsych-2020-100213] [Medline: 32215365]
26. Rehman U, Shahnawaz MG, Khan NH, Kharshiing KD, Khursheed M, Gupta K, et al. Depression, anxiety and stress among Indians in times of Covid-19 lockdown. Community Ment Health J 2021 Jan;57(1):42-48 [FREE Full text] [doi: 10.1007/s10597-020-00664-x] [Medline: 32577997]
27. Roy D, Tripathy S, Kar SK, Sharma N, Verma SK, Kaushal V. Asian J Psychiatr 2020 Jun;51:102083 [FREE Full text] [doi: 10.1016/j.ajp.2020.102083] [Medline: 32283510]
28. Smith L, Jacob L, Yakkundi A, McDermott D, Armstrong NC, Barnett Y, et al. Correlates of symptoms of anxiety and depression and mental wellbeing associated with COVID-19: a cross-sectional study of UK-based respondents. Psychiatry Res 2020 Sep;291:113138 [FREE Full text] [doi: 10.1016/j.psychres.2020.113138] [Medline: 32562931]
29. Shi L, Lu Z, Que J, Huang X, Liu L, Ran M, et al. Prevalence of and risk factors associated with mental health symptoms among the general population in China During the coronavirus disease 2019 pandemic. JAMA Netw Open 2020 Jul 01;3(7):e2014053 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.14053] [Medline: 32609353]
30. Sønderskov KM, Dinesen PT, Santini ZI, Østergaard SD. The depressive state of Denmark during the COVID-19 pandemic. Acta Neuropsychiatr 2020 Aug;32(4):226-228 [FREE Full text] [doi: 10.1017/neu.2020.15] [Medline: 32319879]
31. Tan W, Hao F, McIntyre RS, Jiang L, Jiang X, Zhang L, et al. Is returning to work during the COVID-19 pandemic stressful? A study on immediate mental health status and psychoneuroimmunity prevention measures of Chinese workforce. Brain Behav Immun 2020 Jul;87:84-92 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.055] [Medline: 32335200]
32. Tian F, Li H, Tian S, Yang J, Shao J, Tian C. Psychological symptoms of ordinary Chinese citizens based on SCL-90 during the level I emergency response to COVID-19. Psychiatry Res 2020 Jun;288:112992 [FREE Full text] [doi: 10.1016/j.psychres.2020.112992] [Medline: 32302816]
33. Ustun G. Determining depression and related factors in a society affected by COVID-19 pandemic. Int J Soc Psychiatry 2020 Jun 30:20764020938807 [FREE Full text] [doi: 10.1177/0020764020938807] [Medline: 32605422]
34. Verma S, Mishra A. Depression, anxiety, and stress and socio-demographic correlates among general Indian public during COVID-19. Int J Soc Psychiatry 2020 Dec;66(8):756-762. [doi: 10.1177/0020764020934508] [Medline: 32567466]
35. Wang C, Pan R, Wan X, Tan Y, Xu L, McIntyre RS, et al. A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. Brain Behav Immun 2020 Jul;87:40-48 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.028] [Medline: 32298802]
36. Xiao H, Zhang Y, Kong D, Li S, Yang N. Social capital and sleep quality in individuals who self-isolated for 14 days during the coronavirus disease 2019 (COVID-19) outbreak in January 2020 in China. Med Sci Monit 2020 Mar 20;26:e923921 [FREE Full text] [doi: 10.12659/MSM.923921] [Medline: 32194290]

37. Yuan S, Liao Z, Huang H, Jiang B, Zhang X, Wang Y, et al. Comparison of the indicators of psychological stress in the population of Hubei Province and non-endemic provinces in China during two weeks during the coronavirus disease 2019 (COVID-19) outbreak in February 2020. Med Sci Monit 2020 Apr 15;26:e923767 [FREE Full text] [doi: 10.12659/MSM.923767] [Medline: 32294078]
38. Yuan R, Xu Q, Xia C, Lou C, Xie Z, Ge Q, et al. Psychological status of parents of hospitalized children during the COVID-19 epidemic in China. Psychiatry Res 2020 Jun;288:112953 [FREE Full text] [doi: 10.1016/j.psychres.2020.112953] [Medline: 32302814]
39. Zhang Y, Ma ZF. Impact of the COVID-19 pandemic on mental health and quality of life among local residents in Liaoning Province, China: a cross-sectional study. Int J Environ Res Public Health 2020 Mar 31;17(7):2381 [FREE Full text] [doi: 10.3390/ijerph17072381] [Medline: 32244498]
40. Zhang SX, Wang Y, Rauch A, Wei F. Unprecedented disruption of lives and work: health, distress and life satisfaction of working adults in China one month into the COVID-19 outbreak. Psychiatry Res 2020 Jun;288:112958 [FREE Full text] [doi: 10.1016/j.psychres.2020.112958] [Medline: 32283450]
41. Zhu S, Wu Y, Zhu C, Hong W, Yu Z, Chen Z, et al. The immediate mental health impacts of the COVID-19 pandemic among people with or without quarantine managements. Brain Behav Immun 2020 Jul;87:56-58 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.045] [Medline: 32315758]
42. Lin Y, Liu C, Chiu Y. Google searches for the keywords of "wash hands" predict the speed of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun 2020 Jul;87:30-32 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.020] [Medline: 32283286]
43. Li S, Wang Y, Xue J, Zhao N, Zhu T. The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users. Int J Environ Res Public Health 2020 Mar 19;17(6):2032 [FREE Full text] [doi: 10.3390/ijerph17062032] [Medline: 32204411]

Edited by G Eysenbach; submitted 19.10.20; peer-reviewed by LA Lee, P Banik; comments to author 23.12.20; revised version received 29.12.20; accepted 16.01.21; published 10.02.21.

Please cite as:
Lin YH, Chen CY, Wu SI
Efficiency and Quality of Data Collection Among Public Mental Health Surveys Conducted During the COVID-19 Pandemic: Systematic Review
J Med Internet Res 2021;23(2):e25118
URL: http://www.jmir.org/2021/2/e25118/
doi:10.2196/25118
PMID:33481754

Original Paper

Fast and Accurate Detection of COVID-19 Along With 14 Other Chest Pathologies Using a Multi-Level Classification: Algorithm Development and Validation Study

Saleh Albahli[1,2*], BSc, PhD; Ghulam Nabi Ahmad Hassan Yar[3*], BS, MS

[1]Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia

[2]Department of Computer Science, Kent State University, Kent, OH, United States

[3]Depratment of Electrical and Computer Engineering, Air University, Islamabad, Pakistan

[*]all authors contributed equally

Corresponding Author:
Saleh Albahli, BSc, PhD
Department of Information Technology
College of Computer
Qassim University
Buraydah, 51452
Saudi Arabia
Phone: 966 163012604
Email: salbahli@qu.edu.sa

Abstract

Background: COVID-19 has spread very rapidly, and it is important to build a system that can detect it in order to help an overwhelmed health care system. Many research studies on chest diseases rely on the strengths of deep learning techniques. Although some of these studies used state-of-the-art techniques and were able to deliver promising results, these techniques are not very useful if they can detect only one type of disease without detecting the others.

Objective: The main objective of this study was to achieve a fast and more accurate diagnosis of COVID-19. This study proposes a diagnostic technique that classifies COVID-19 x-ray images from normal x-ray images and those specific to 14 other chest diseases.

Methods: In this paper, we propose a novel, multilevel pipeline, based on deep learning models, to detect COVID-19 along with other chest diseases based on x-ray images. This pipeline reduces the burden of a single network to classify a large number of classes. The deep learning models used in this study were pretrained on the ImageNet dataset, and transfer learning was used for fast training. The lungs and heart were segmented from the whole x-ray images and passed onto the first classifier that checks whether the x-ray is normal, COVID-19 affected, or characteristic of another chest disease. If it is neither a COVID-19 x-ray image nor a normal one, then the second classifier comes into action and classifies the image as one of the other 14 diseases.

Results: We show how our model uses state-of-the-art deep neural networks to achieve classification accuracy for COVID-19 along with 14 other chest diseases and normal cases based on x-ray images, which is competitive with currently used state-of-the-art models. Due to the lack of data in some classes such as COVID-19, we applied 10-fold cross-validation through the ResNet50 model. Our classification technique thus achieved an average training accuracy of 96.04% and test accuracy of 92.52% for the first level of classification (ie, 3 classes). For the second level of classification (ie, 14 classes), our technique achieved a maximum training accuracy of 88.52% and test accuracy of 66.634% by using ResNet50. We also found that when all the 16 classes were classified at once, the overall accuracy for COVID-19 detection decreased, which in the case of ResNet50 was 88.92% for training data and 71.905% for test data.

Conclusions: Our proposed pipeline can detect COVID-19 with a higher accuracy along with detecting 14 other chest diseases based on x-ray images. This is achieved by dividing the classification task into multiple steps rather than classifying them collectively.

(J Med Internet Res 2021;23(2):e23693) doi:10.2196/23693

KEYWORDS

COVID-19; chest x-ray; convolutional neural network; data augmentation; biomedical imaging; automatic detection

Introduction

Background

The COVID-19 pandemic has been causing significant health concerns since 2019. Symptoms of the disease include fever, cough, headache, and severe respiratory complications, which can subsequently lead to death. When this disease first started to spread in December 2019, numerous unknown facts were reported in Wuhan, China, where the first outbreak occurred [1]. By early January 2020, the government of China and the World Health Organization recognized SARS-CoV-2, the novel coronavirus known to cause COVID-19, as a pathogenic virus that belongs to the same family (Coronaviridae) as the virus known to cause severe acute respiratory syndrome (SARS). A SARS outbreak was previously reported in China in 2002-2003 [2].

Medical x-rays (short for x-radiation) are a form of visible light rays but with higher energy that penetrate the body to generate images of tissues and structures within the body, including bones, chest, and teeth. X-ray imaging is a very effective diagnostic tool and has been used for several decades by specialists to detect fractures, certain tumors, pneumonia, and dental problems [3]. In advanced cases, computed tomography (CT) can be used to produce a series of body images, which is later assembled into a 3D x-ray image that is processed by a computer. However, the traditional x-ray is a lot faster, easier, cheaper, and less harmful than a CT scan [4].

Research has shown that deep learning can be used to make predictions based on medical images by extracting characteristic features, including the shape and spatial rotation, from the images. Convolutional neural networks (CNNs) have played a very vital role in feature extraction and learning patterns that enable prediction. For example, a CNN is used to improve extraction high-speed video-endoscopy when the training data is very limited [5]. Advancements in image processing tools have brought about a radical change in the current techniques for the detection of pulmonary diseases. Researchers are employing traditional computer vision as well as deep learning algorithms to achieve satisfactory performance [3]. Several primary benefits are strongly correlated with the advancement of radiographic image classification tools. For example, in rural areas, owing to a shortage of doctors and places where doctors cannot be reached, such tools can prove useful. Once these tools become pervasive in the health care industry, radiologists, clinic practitioners, and even patients may utilize radiographic image classification tools to monitor and treat several diseases. As a result, this can reduce the burden on radiologists all over the world, by abolishing the requirement to examine every x-ray image for anomalies. Instead, the doctors will only need to focus on the patients whose x-ray images are flagged by this tool. The use of such tools can also eliminate the subjective opinion of doctors, increase the speed of early diagnosis of disease, and identify the minor details that may be overlooked by the human eye in some cases.

For this study, owing to computational restraints, we did not build a model from scratch, as such models require extremely high-end computers. Rather, we used CNN as a class of deep neural networks to propose a model to classify COVID-19 x-ray images from x-ray images of a wide range of chest diseases. Although the x-ray images of the other diseases are inadequate for proper training and to achieve state-of-the-art results, we generalized the data by considering data augmentation. This mainly rescales the x-ray images and flips them horizontally, in addition to a few other functionalities such as shift range, zooming, and rotation.

The strength of this study is that it classifies x-ray images at two different stages. The first stage involves enhancing the model to detect COVID-19–specific x-ray images at a faster speed than x-ray images of other chest diseases. This will result in a significant increase in the classification speed. Thus, considering a part of a dataset of chest x-ray (CXR) images for the analysis will result in low-quality output and unsatisfactory diagnoses. Accordingly, if the case is not classified as “normal” or “COVID-19” at this stage, then the classification is continued to the second stage, which involves classification for 14 other chest and related conditions (ie, atelectasis, cardiomegaly, effusion, infiltration, mass, nodule, pneumonia, pneumothorax, consolidation, edema, emphysema, fibrosis, pleural, and hernia). This also saves processing power if the x-ray image has been classified as “normal” or “COVID-19” in the first stage itself. To further enhance the accuracy of detection, we used UNet to complete lung and heart segmentation. Because we used a pretrained model, we were able to independently train 5 different models for each stage. Models with the best training and test accuracy were then selected for further analyses.

Based on our findings, we found that ResNet50 is the best model for classification in both scenarios: classifying 3 classes and 14 classes. Moreover, image segmentation helps in increasing the classification accuracy by up to 5%. We also trained a model for all 16 classes and found that classifying for a large number of classes significantly reduces the overall accuracy of the model.

The main contributions of this study are as follows:

1. Introduction of new classification pipeline for more accurate, automated classification in case of a large number of classes, primarily to increase the accuracy of a specific class.
2. Use of augmentation and semantic segmentation to increase accuracy of the model.
3. Comparison between different deep learning models on the basis of classification in cases of small and large number of classes.

In this paper, we first review previous studies that used deep neural networks for the detection of COVID-19 and other chest diseases. Then, we discuss the datasets used for our experiments as well as the study methodology, including data preprocessing, data segmentation, and the setup for classification of the models.

Finally, we present the results and analyses on the basis of the models and dataset available.

Previous Work

Recently, with the rapid development of artificial intelligence, an increasing number of researchers have begun to pay attention to intelligent, deep learning−based diagnostic techniques. Some of them have achieved significantly prominent results. In this section, we first review the current, state-of-the-art techniques concerning the application of artificial intelligence to chest diseases in general, and then, we discuss the literature related to COVID-19 detection using deep neural networks.

Detection of Chest Diseases Based on CXR Images by Using Deep Neural Networks

Sivasamy and Subashini [6] used a Keras framework to classify CXR images to predict lung diseases and reported an accuracy of 86.14%. The accuracy of the model improved as the number of epochs for training was increased. Wang et al [7] used pixel-wise annotated digital reconstructed radiograph data to train an unsupervised multiorgan segmentation model based on x-ray images. In this case, the gaps in nodules annotated directly on 2D x-ray images are quite challenging and time-consuming due to the projective nature of x-ray imaging. Rajpurkar et al [8] proposed a binary classifier for the detection of pneumonia from frontal-view CXR images that achieved an f1 score of 0.435. Salehinejad et al [9] used a Deep Convolutional Generative Adversarial Network (DCGAN) tailored model designed for training with x-ray images wherein a generator is trained to generate artificial CXR images. Their model obtained its best accuracy when trained on an augmented dataset with DCGAN-synthesized CXRs to balance the imbalanced real dataset (D3). Chandra and Verma [10] used 5 different models to identify pneumonia and reported 95.631% as the best accuracy. The model is limited to analyzing only nonrigid, deformable, registration-driven automatically lung regions and segmented region of interest–confined feature extraction. Previous studies using state-of-the-art techniques have achieved effective results with one or two cardiothoracic diseases, but these techniques could lead to misclassification.

A few techniques have targeted all 14 classes of chest diseases. Wang et al [11] presented the largest publicly available dataset of CXR images, which has provided a new dimension to the research community. They achieved promising results using a deep CNN and suggest that this dataset could be further extended by using more disease labels. Smit et al [12] proposed a deep learning−based technique to identify the 14 underlying chest diseases. They trained the model to input a single-view chest radiograph and output the probability of each of the 14 observations. Several models were trained to identify the one with the best accuracy. They used DenseNet121 for their research and found that it yielded the best accuracy, but it was limited to the CheXpert dataset and liable to overfitting. A pretrained DenseNet121 model and feature extraction techniques were used for accurate identification of 14 thoracic diseases in the study by Ho and Gwak [13].

Detection of COVID-19 Cases Based on CXR Images by Using Deep Neural Networks

There are several state-of-the-art studies on deep learning and machine learning models for COVID-19 diagnosis. A study by Apostolopoulos and Mpesiana [14] took advantage of CNNs for the automatic detection of COVID-19 by using CXR images. They adopted transfer learning to solve for the small image dataset challenge. Their COVID-19 dataset consisted of 224 sample medical images. Despite the size limitation, their results showed effective automatic detection of COVID-19−related diseases. Abbas et al [15] used the CNN-based DeTraC framework. They also used transfer learning to achieve the best performance. This model achieved 95.12% accuracy and 97.91% sensitivity. Chen et al [16] provided a prediction of patients with or without COVID-19 by using the UNet++ based segmentation model. Narin et al [17] classified CXR images using the ResNet50 model and obtained the highest classification performance with 98% accuracy, using a dataset comprising only 50 COVID-19 and 50 normal samples. Li et al [18] also used a ResNet50 model with a dataset comprising 468 COVID-19 samples, 1551 community-acquired pneumonia samples, and 1445 non-pneumonia samples; this model achieved 90% sensitivity. Using deep learning approaches to extract and transform features, Li et al proved their model's efficacy in COVID-19 diagnosis [18]. Furthermore, Sethy and Behera [19] used deep learning to extract deep features from x-ray images and then used state vector machine to classify them into COVID-19–positive and COVID-19–negative classes; they achieved an accuracy of 95%. Hemdan et al [20] used transfer learning and fine-tuning on state-of-the-art networks like VGG and ResNetV2 to classify COVID-19–positive and COVID-19–negative x-ray images; they achieved an accuracy of 90%. Wang et al [21] proposed the M-inception model, a variant of the inception model. They detected only COVID-19 CT images from all available images and achieved an accuracy of 82%. Table 1 presents a comparison of previously studies models using radiographic imaging classification for COVID-19 cases, normal cases, and other chest diseases.

Table 1. Comparison of models detecting COVID-19 cases, normal cases, and other chest diseases based on medical images (data derived from [22]).

Reference	Medical image	Disease detected, n			Accuracy (%)	Methodology	Gaps in classification
		COVID-19	Normal	Other chest diseases			
Apostolopoulos and Mpesiana [14]	X-ray	224	504	700	93	Used transfer learning on VGG19. MobileNetV2, Inception, Xception, and InceptionResNetV2	Used only 3 classes: COVID-19, pneumonia, and other
Wang et al [23]	X-ray	53	8066	5526	92	Introduced COVID-Net—the first open-source COVID-19 detection system	Used only 3 classes: COVID-19, pneumonia, and normal
Narin et al [17]	X-ray	50	50	N/A[a]	98	Used 5 pretrained networks and applied 3 binary classifications for 4 classes of chest x-rays	Used only 3 classes: normal, COVID-19, viral and bacterial pneumonia
Brunese et al [22]	X-ray	250	3520	2753	97	Defined 2 models based on VGG16: one to classify affected x-ray images from healthy ones and the other to classify COVID-19 from affected x-ray images. Then, they localized the affected areas.	Although they used x-ray images of most diseases, they used only 3 classes: COVID-19, healthy, and disease
Song et al [24]	CT[b]	777	708	N/A	86	Proposed DRE-Net and compared its performance with VGG-16, DenseNet, and ResNet	Used only 3 classes: COVID-19, bacterial pneumonia, and healthy
Zheng et al [25]	CT	313	229	N/A	90	Proposed DeCoVNet for classification	Used only 2 classes: COVID-19–positive and COVID-19–negative
Xu et al [26]	X-ray	219	175	224	86	Proposed ResNet-18 based CNN[c] network	Used only 3 classes: COVID-19, Influenza-A viral pneumonia, and normal
Ozturk et al [27]	X-ray	250	1000	500	92	Proposed DarkCovidNet	Used only 3 classes: COVID-19, pneumonia, and no findings
Ardakani et al [28]	CT	510	N/A	510	99	Used 10 CNN networks (ie, AlexNet and ResNet-101) for classification of 2 classes	Classified COVID-19 class from non–COVID-19 class
Li et al [18]	CT	1296	1325	1735	96	Proposed COV-Net for classifying 3 classes	Used only 3 classes: COVID-19, community-acquired pneumonia, and non-pneumonia
Abbas et al [15]	X-ray	105	80	11	95.12	Proposed DeTrac-ResNet18 CNN that uses Decompose, Transfer, and Compose architecture	Used only 3 classes: normal, COVID-19, and SARS
Chen et al [16]	CT	51	N/A	55	95.24	Used UNet++ along with Keras for segmentation and COVID-19 detection	Used only binary classification for COVID-19 detection

[a]N/A: not applicable.

[b]CT: computed tomography.

[c]CNN: convolutional neural network.

Methods

Dataset

The first step involved preprocessing of the data, which includes segmentation of the lungs and the heart from the whole image, as an x-ray image contains many unnecessary details. To perform this segmentation task, we trained the UNet model on segmented CXR data obtained by the Japanese Society of Radiological Technology, which were downloaded from their official website [29], and their corresponding masks, which were downloaded from the SCR database [30]. This dataset contains 247 images. For classification purposes, data for COVID-19 was collected from Cohen et al's COVID Chest X-ray dataset [31]. This dataset contains x-ray images of many other diseases. Furthermore, x-ray images from the datasets were separated using the available metadata file. Data for the other 14 chest diseases were provided by the National Institute of Health (NIH) and can be downloaded from the NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories [32]. Data available on the NIH Clinical Center website contains 112,120 images, belonging to 15 classes, which include 14 disease classes and 1 normal class—all of which were extracted through the available metadata file. The number of images per class is presented in Table 2.

Table 2. Number of images per class in the National Institute of Health Chest X-ray Dataset of 14 Common Thorax Disease Categories [32].

Model and class	Training set, n	Testing set, n
Model 1		
COVID-19	455	22
Normal	1995	405
Other	4600	730
Model 2		
Atelectasis	200	100
Cardiomegaly	200	100
Consolidation	200	100
Edema	200	100
Effusion	200	100
Emphysema	200	100
Fibrosis	200	100
Hernia	150	100
Infiltration	200	100
Mass	200	100
Nodule	200	100
Pleural thickening	200	100
Pneumonia	200	100
Pneumothorax	200	100

The data were randomly split into training and testing sets, as there were very few data related to COVID-19. The idea was to keep the training set as large as possible given the small number of images present. Image augmentation compensated for the lack of data. This was not an issue for model 2 images. For model 1, however, the lack of data can cause a change in testing accuracy. To compensate for this issue, we also applied data augmentation while testing.

Data Preprocessing

Every x-ray image has a different contrast and illumination as they are taken under different lighting conditions. Therefore, in the first step of preprocessing, histogram equalization was applied. CXR images also contain unnecessary details, such as the collarbone, shoulders, neck, and torso region. To remove these unnecessary details, lungs and heart segmentation were applied. For this purpose, the UNet segmentation model was trained on images from the Japanese Society of Radiological Technology with their corresponding masks. The architecture of the UNet model is shown in Table 3. The input image size fed to the network was 256×256×3. The contraction part acts as an encoder that extracts the context from the image using downsampling through the max-pooling layer. The expansive path acts as a decoder that precisely localizes the segmentation part using transpose convolution layers. It is an end-to-end, fully connected network and does not contain any dense layers. It also restores the image through upsampling.

Table 3. Architecture of UNet model.

Path, layer, and type		Kernel size	Filters
1	Input Layer	N/A[a]	N/A
Contraction Path			
2	Convolution	3×3	16
3	Dropout (0.1)	N/A	N/A
4	Convolution	3×3	16
5	MaxPooling	2×2	1
6	Convolution	3×3	32
7	Dropout (0.1)	N/A	N/A
8	Convolution	3×3	32
9	MaxPooling	2×2	1
10	Convolution	3×3	64
11	Dropout (0.2)	N/A	N/A
12	Convolution	3×3	64
13	MaxPooling	2×2	1
14	Convolution	3×3	128
15	Dropout (0.2)	N/A	N/A
16	Convolution	3×3	128
17	MaxPooling	2×2	1
18	Convolution	3×3	256
19	Dropout (0.3)	N/A	N/A
20	Convolution	3×3	256
Expansive Path			
21	Transposed convolution	2×2	128
22	Concatenate (21, 16)	N/A	N/A
23	Convolution	3×3	128
24	Dropout (0.2)	N/A	N/A
25	Convolution	3×3	128
26	Transposed convolution	2×2	64
27	Concatenate (26, 12)	N/A	N/A
28	Convolution	3×3	64
29	Dropout (0.2)	N/A	N/A
30	Convolution	3×3	64
31	Transposed convolution	2×2	32
32	Concatenate (31, 8)	N/A	N/A
33	Convolution	3×3	32
34	Dropout (0.1)	N/A	N/A
35	Convolution	3×3	32
36	Transposed convolution	2×2	16
37	Concatenate (36, 4)	N/A	N/A
38	Convolution	3×3	16
39	Dropout (0.1)	N/A	N/A
40	Convolution	3×3	16

Path, layer, and type		Kernel size	Filters
41	Convolution (Sigmoid)	1×1	1

[a]N/A: not applicable.

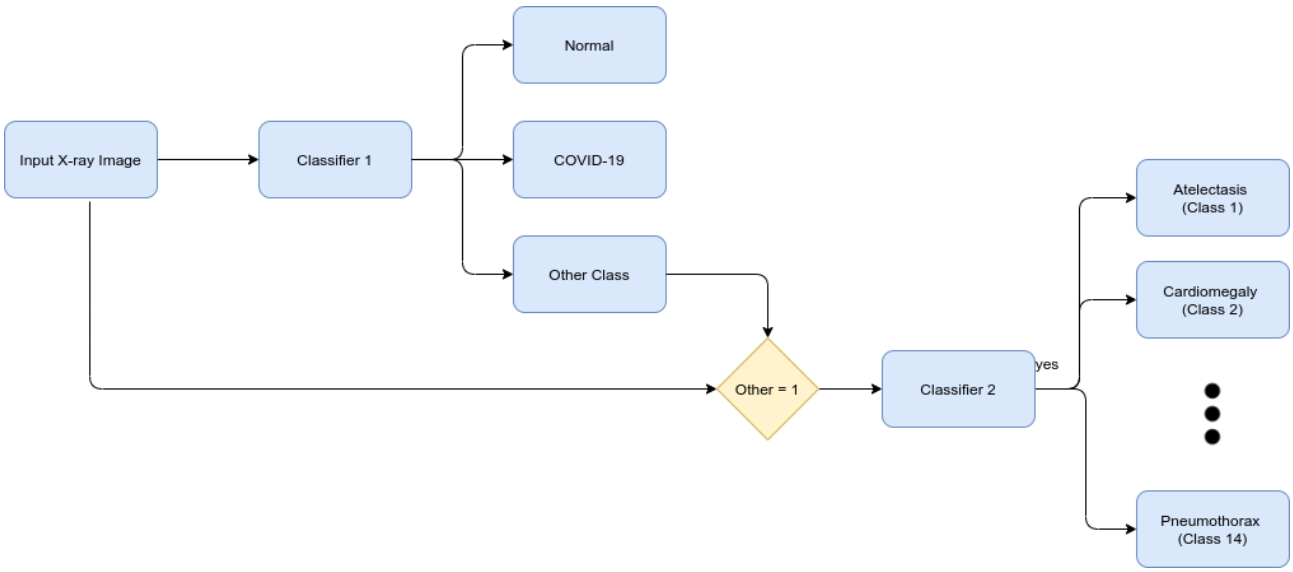
Data Augmentation

Before feeding the data to the network, image augmentation was applied to tackle the problem of fewer data, that is, in the case of COVID-19. For applying augmentation, the rotation range was set to 90°, the horizontal flip was set to true, and the vertical flip was also set to true. For each iteration, the image data generator used a different transformation of the original images. In the case of COVID-19, we had 445 input images and 20 iterations; therefore, the data generator used 8900 images for training in this case.

Classification Models

The main objective of this study was to classify COVID-19 x-ray images from normal x-ray images and those of 14 other chest diseases. When a single model is trained for classifying 16 different classes, its accuracy tends to decrease, and in the case of COVID-19 detection, that is not acceptable. To solve this problem, a new pipeline was formed, which is illustrated in Figure 1. Two models were trained. The first model was trained to classify 3 classes: COVID-19, normal, and some other disease. The second model was trained to classify the 14 other chest and related diseases. Both models were trained separately. To automate the process, if the first model classified the x-ray as "some other disease," then the second model was called to further classify the disease as one of 14 other chest diseases, using a simple "IF" condition. This architecture makes the classification process easy, as there are fewer features that need to be classified at the first stage.

Figure 1. Proposed pipeline of classification.



Classifier 1 only needs to learn how to distinguish COVID-19 and normal x-ray images from those of all the other 14 chest diseases. The rule is simple: the fewer the classes, the fewer features there are to learn and distinguish, and the greater the accuracy. This is critical because the classification of COVID-19 is much more important than that of other diseases during the ongoing pandemic. Finally, the burden of classifying the other 14 x-ray diseases falls on classifier 2, which now has 14 classes to classify instead of 16. Furthermore, the 2 most important classes have already been classified by classifier 1. Moreover, to support the statement that accuracy indeed decreases when classifying into 16 classes, a third model was trained for classification into all 16 classes.

For classification purposes, the following 5 models were trained for both classifications:

NasNetLarge, Xception, InceptionV3, InceptionResNetV2, and ResNet50.

NasNetLarge was proposed by "Google Brain" in 2018 [33]. It has two types of architectures—CIFAR10 and ImageNet. CIFAR10 architecture has N number of normal cells and one reduction cell repeating after each other; in the end, it has the SoftMax function. ImageNet has two strides of convolutional layers with a 3×3 kernel size at the start, followed by two reduction cells; thereafter, it has the same architecture as CIFAR10.

Xception was proposed by Google in 2017 [34]. It consists of one entry flow, eight middle flow, and one exit flow. Entry flow consists of convolutional and max-pooling layers with ReLU as the activation function. The middle flow consists of only convolutional layers with the ReLU activation function. Exit flow consists of convolutional, max pooling, and global average pooling layers with the ReLU activation function; in the end, it has fully connected layers for classification.

InceptionV3 was proposed by Google in 2015 [35]. The basic architecture of the model is the same, as it consists of convolutional and pooling layers; in addition, it has three

inception architectures as proposed previously [35]. Finally, at the end, it has the logistic and SoftMax function for classification into 1000 classes.

InceptionResNetV2 was proposed by Google in 2016 [36]. It has the proposed inception and reduction blocks at the start, and in the end, it has a pooling layer and dropout layer to prevent overfitting. It classifies using the SoftMax function.

ResNet50 was proposed by Microsoft in 2015 [37]. It takes residual learning as a building block and consists of convolutional layers with an average pooling layer at the end.

Models were taken from Keras library in Python, which were initialized with ImageNet weights. These models can classify 1000 classes, but we only needed to classify 3, 14, and 16 classes for classifier 1, classifier 2, and classifier 3, respectively. Therefore, these models were fine-tuned, and additional layers were added. Table 4 shows the fine-tuning layers added at the end of each pretrained model. The input image size given to the models was 331×331×3.

Table 4. Fine-tuning layers for classifier 1, classifier 2, and classifier 3.

Type	Classifier 1		Classifier 2		Classifier 3	
	Output	Kernel	Output	Kernel	Output	Kernel
Average pooling	2048	2×2	2048	2×2	2048	2×2
Flatten	8192	N/A[a]	8192	N/A	8192	N/A
Dense	1024	N/A	1024	N/A	1024	N/A
Dropout (0.5)	1024	N/A	1024	N/A	1024	N/A
Dense	1024	N/A	1024	N/A	1024	N/A
Dropout (0.5)	1024	N/A	1024	N/A	1024	N/A
Dense	3	N/A	1024	N/A	16	N/A

[a]N/A: not applicable.

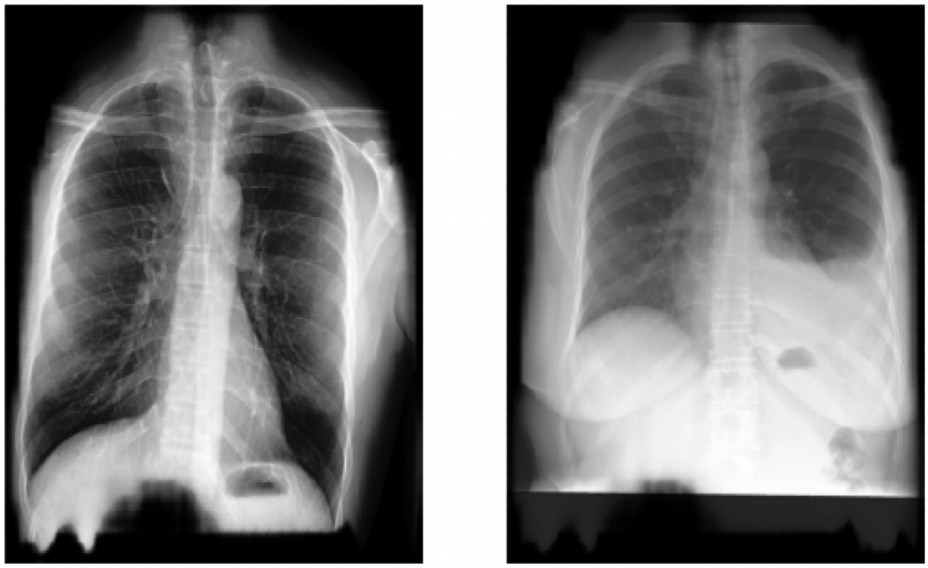
All the models explained in the Methods section were trained and tested on Google Colab with 12 GB of RAM and GPU (graphics processing unit) assigned by Google Colab.

Results

Initially, the UNet model was trained for segmentation of lungs and heart. After Training UNet, the model had a training loss of 16.75%, training accuracy of 87.13%, validation loss of 12.27%, and validation accuracy of 89.64%.

Figure 2 shows some sample segmented CXR images. With image segmentation, we achieved up to 5% increase in the accuracy of our models.

Figure 2. Sample segmented chest x-ray images.



After image segmentation was completed and new training data were obtained, each training model was trained for 20 epochs with a batch size of 8. The accuracy obtained from training is shown in Table 5. The table shows that the maximum accuracy for classifying the 3 classes, including COVID-19 was achieved by using ResNet50 followed by NasNetLarge. These two models yielded accuracy that competes with that of the available state-of-the-art models for COVID-19 prediction.

Classifier 2 did not show promising results in classifying the 14 other diseases. The main reasons for this were the large number of classes and the continued overfitting of the model. The maximum test accuracy achieved was 65.63% with ResNet50 followed by 61.47% with NasNetLarge.

As described, our proposed model pipeline helps to increase the accuracy of COVID-19 diagnosis when classifying 16 classes. Table 5 shows that when the 16 classes were combined and classified, the detection accuracy decreases. In all the cases except in the case of NASNetLarge and ResNet50 models, the test accuracy decreased when classifying 16 classes. Moreover, in the case of the NASNetLarge model, the increase in accuracy is not very notable. The maximum test accuracy was achieved with ResNet50, with an average of 71.905% over 10-fold cross-validation.

Table 5. Average training, validation, and test accuracy achieved by different models through a 10-fold cross-validation.

Model and classifier	Training accuracy (%)	Training loss (%)	Validation accuracy (%)	Validation loss (%)	Test accuracy (%)	Test loss (%)
NasNetLarge						
1st	91.80	33.78	91.25	31.32	89.66	32.028
2nd	84.67	52.45	61.22	127.83	61.47	127.99
Combined	79.72	68.36	63.68	123.42	63.58	121.39
Xception						
1st	88.12	29.27	87.33	36.27	86.58	35.91
2nd	90.70	48.73	61.88	133.04	61.08	133.92
Combined	29.88	208.89	22.28	397.08	47.75	301.39
InceptionV3						
1st	65.87	69.25	63.43	57.38	63.19	5.233
2nd	83.91	56.46	53.57	142.22	53.75	139.97
Combined	65.52	110.33	38.31	176.24	38.83	174.97
InceptionResNetV2						
1st	65.10	80.32	62.46	76.35	63.19	75.79
2nd	83.37	81.30	54.08	197.66	53.75	197.07
Combined	54.45	134.20	33.84	200.61	33.97	200.26
ResNet50						
1st	96.32	9.84	94.16	23.09	92.52	20.32
2nd	87.83	35.85	67.55	105.63	65.63	108.24
Combined	88.92	26.16	73.14	87.05	71.91	88.95

The results obtained by our proposed approaches compete with that of state-of-the-art methods (shown in Table 1). Graphs illustrating the training and validation accuracy and loss for classifiers 1, 2, and 3 are shown in Figures 3, 4, and 5, respectively. To further evaluate the results, the AUC (area under the curve), sensitivity, and specificity results for all the networks were studied (Table 6). We found that ResNet50 achieved the maximum AUC, sensitivity, and specificity scores compared to any other model.

Figure 3. Graphs illustrating training and validation accuracy (left) and loss (right) over epochs for different models of classifier 1.

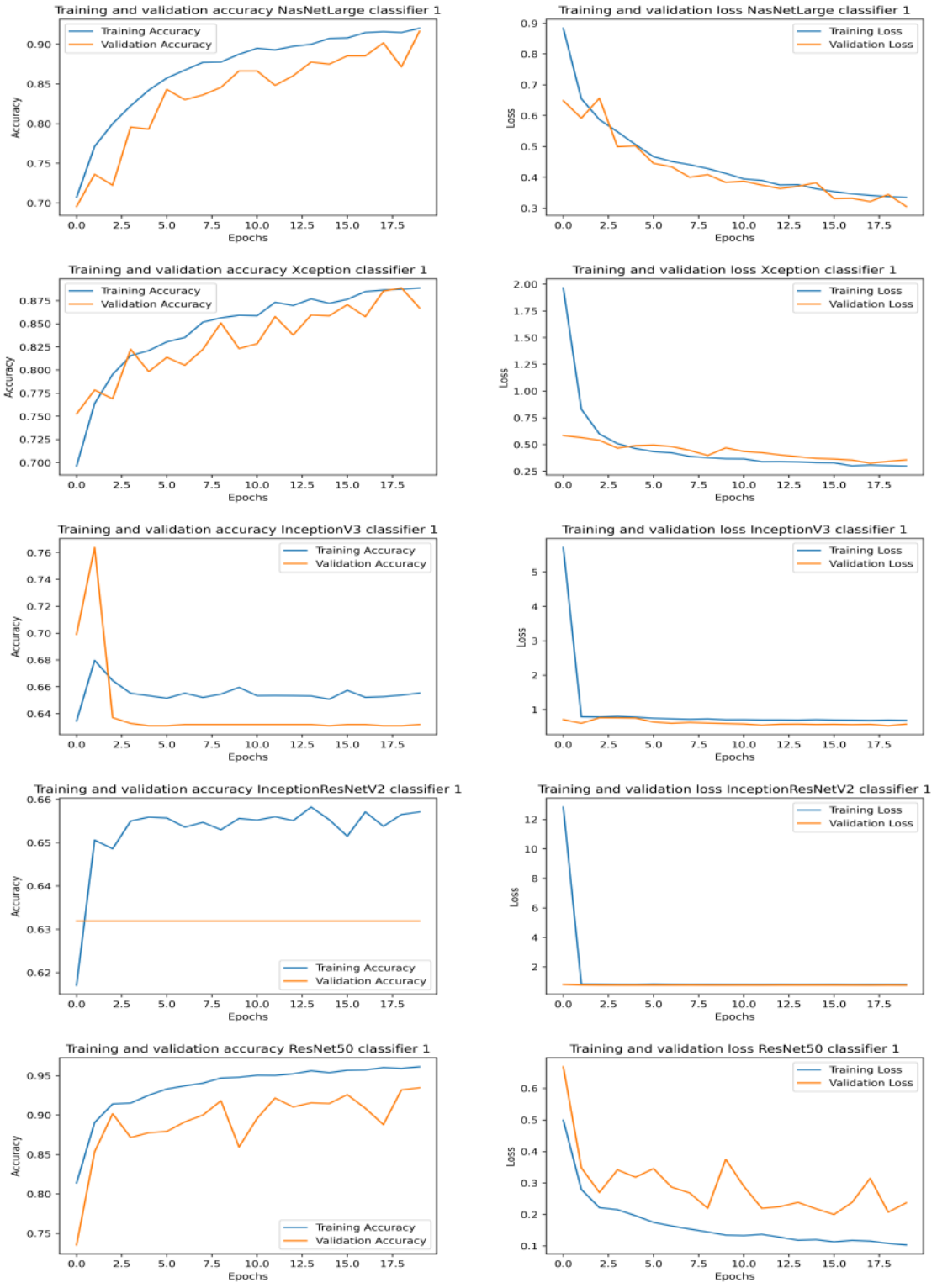


Figure 4. Graphs illustrating training and validation accuracy (left) and loss (right) over epochs for different models of classifier 2.

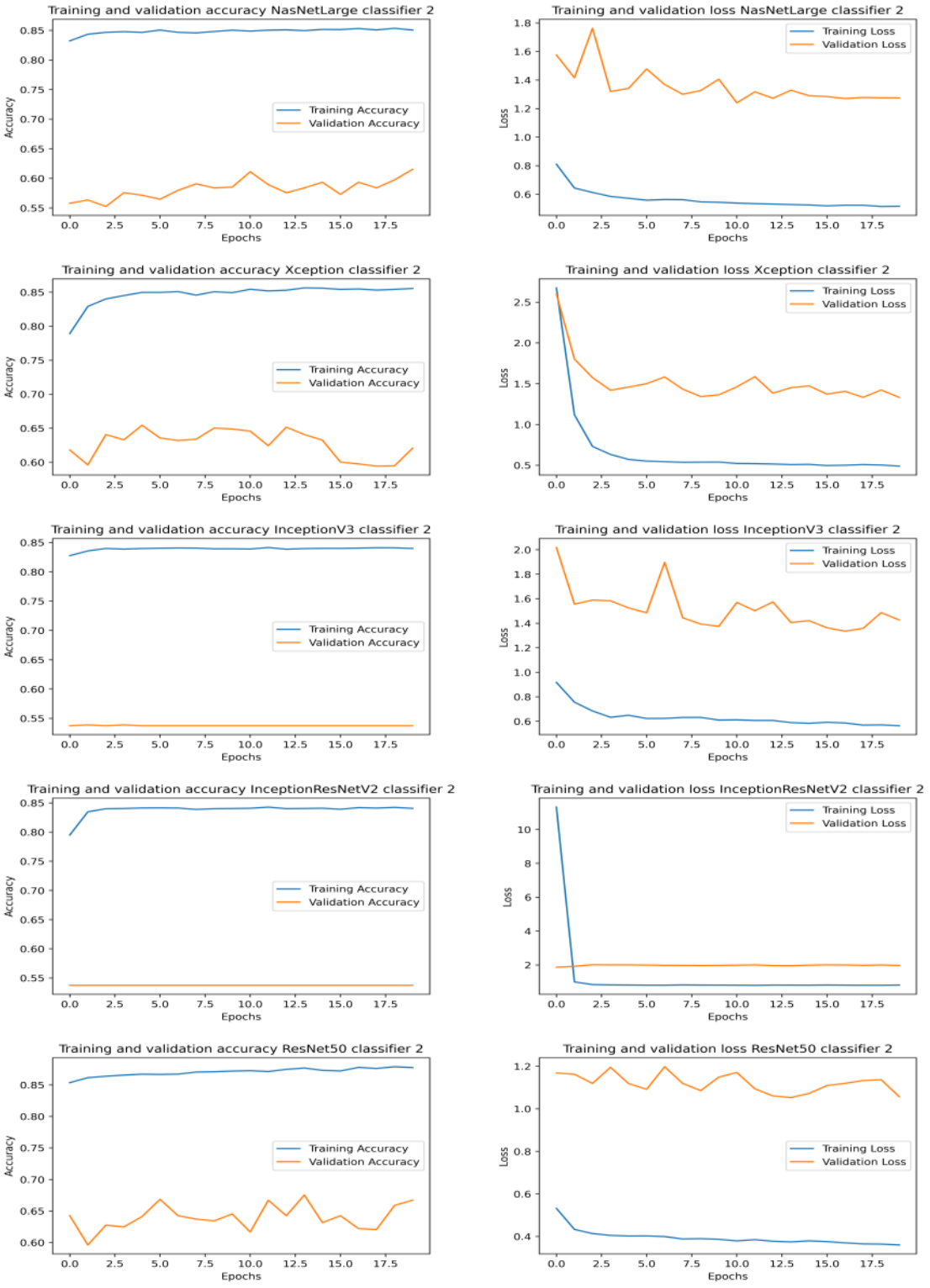


Figure 5. Graphs illustrating training and validation accuracy (left) and loss (right) over epochs for different models of classifier 3.

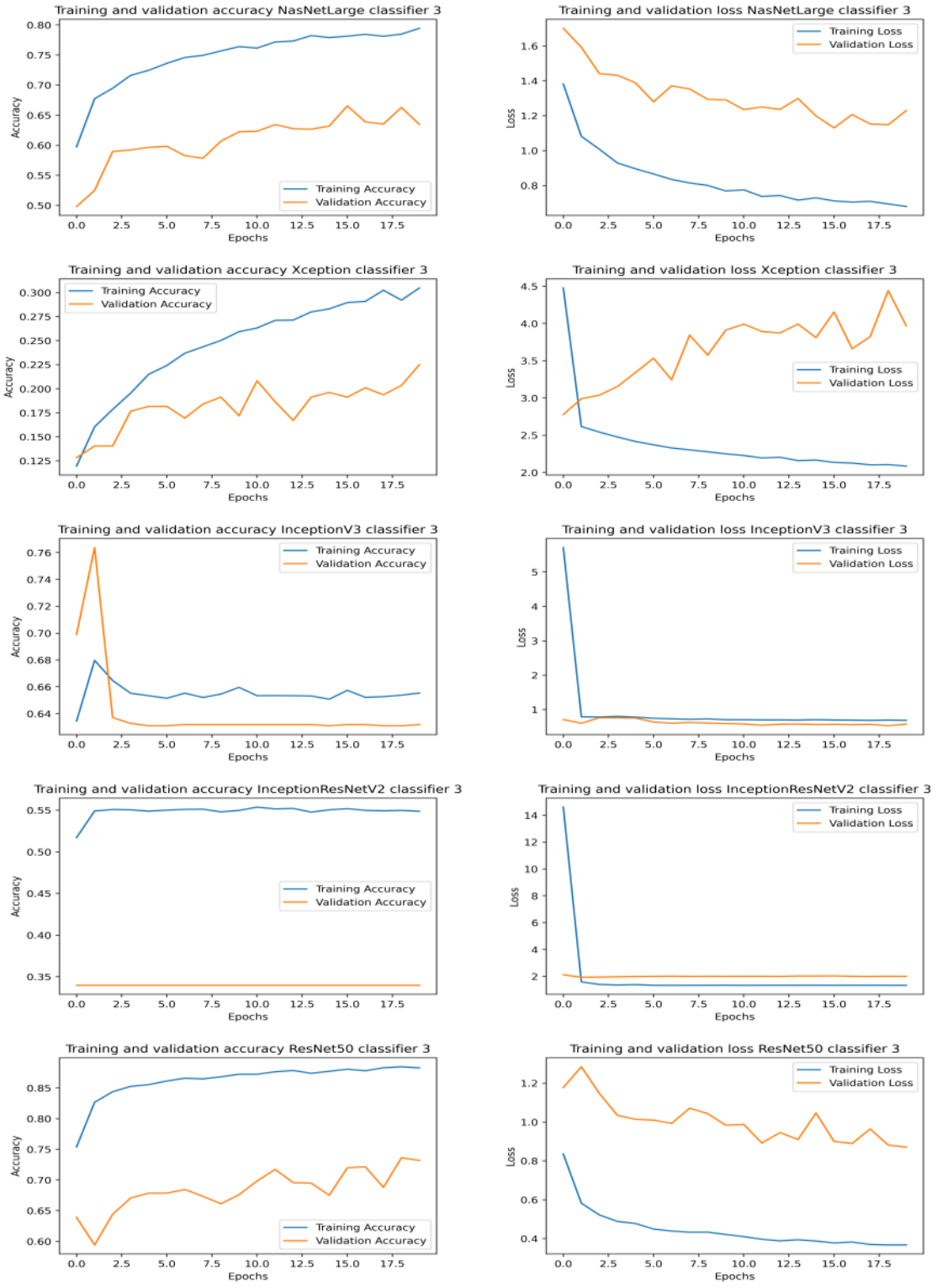


Table 6. Average training, validation, and test accuracy achieved by different models through the 10-fold cross-validation.

Model and classifier	AUC[a] (%)	Sensitivity (%)	Specificity (%)
NasNetLarge			
1st	97.61	90.73	93.42
2nd	82.15	75.33	81.1
Combined	93.88	90.15	93.85
Xception			
1st	95.9	88.64	91.78
2nd	92.25	87.38	81.12
Combined	83.19	76.09	80.64
InceptionV3			
1st	89.28	83.2	85.51
2nd	91.69	79.53	81.25
Combined	89.85	83.29	93.7
InceptionResNetV2			
1st	80.65	74.88	77.92
2nd	85.41	79.21	81.25
Combined	85.44	82.69	93.75
ResNet50			
1st	98.73	93.14	95.22
2nd	94.6	85.64	81.03
Combined	96.9	93.4	93.72

[a]Area under the curve.

Discussion

Principal Findings

In this study, we classified normal cases, COVID-19 cases, and 14 other chest diseases based on CXR images. We proposed a novel, multiclass method for this purpose and used models that were pretrained on ImageNet dataset to save training time and resources. Our multilevel approach resulted in an increase in the classification accuracy. We found that ResNet50 was the best model for classification, yielding the highest accuracy.

Future Suggestions

This study tried to cover most aspects of detection of chest diseases, but there is still work to be done. Most importantly, there is a need for more data for patients with COVID-19, which could help improve the accuracy of the model. At present, there is a significant difference in the number of images per class for the first level of classification.

This model can help in the first level of classification to determine whether the person has COVID-19 or some other chest disease, as x-rays are easier and less expensive than other forms of radiographic imaging and can help determine the severity of the disease. Although disease severity was not within the scope of this study, future work in detecting the severity of the disease can also be an important improvement in the already-existing model. In addition, techniques such as the Grad-Cam algorithm can be used to visualize the features in radiographic images affecting the algorithm and to determine disease severity. This algorithm will highlight which features help the algorithm with the classification and which features likely mislead the algorithm. This algorithm might also be the key to investigating the low accuracy of the level-2 classifier and can help improve its accuracy.

Conclusions

Deep learning has played a major role in medical image analysis and feature extraction, which are applied to the detection of a wide range of chest diseases. CNN architectures are popular for their ability to learn mid- and high-level image representations and to make predictions. Detecting the presence, or absence, of COVID-19 in a patient is insufficient without addressing other chest diseases. However, a deep learning system that is trained to classify a large number of classes—16 in our case—has less accuracy. This work aimed to deal effectively with this new pipeline to help with a first-level differential diagnosis of COVID-19 from other chest diseases. Subsequently, we applied further enhancement to detect other chest diseases in order to tackle multi-class chest classification in the detection of anomalies on x-ray images. This approach yielded satisfactory results.

Thus, we showed how our proposed models use state-of-the-art deep neural networks to classify 16 cardiothoracic diseases by training the models based on x-ray images in the database. Image

segmentation was applied to remove unnecessary details, and both classifiers were independently trained on segmented data. However, our model can classify not only COVID-19 but also 14 other chest diseases, as well as normal x-ray images, with satisfactory accuracy as compared with previous studies.

Acknowledgments

We would like to thank the Deanship of Scientific Research, Qassim University, for funding the publication of this project.

Conflicts of Interest

None declared.

References

1. Coronavirus disease 2019 (COVID-19) Situation Report – 94. World Health Organization. 2019. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200423-sitrep-94-covid-19.pdf [accessed 2021-02-04]
2. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. The Lancet Respiratory Medicine 2020 May;8(5):475-481. [doi: 10.1016/s2213-2600(20)30079-5]
3. Albahli S. A deep neural network to distinguish COVID-19 from other chest diseases using x-ray images. Curr Med Imaging. Epub ahead of print posted online on June 4, 2020 2020. [doi: 10.2174/1573405616666200604163954] [Medline: 32496988]
4. Haglin JM, Jimenez G, Eltorai AEM. Artificial neural networks in medicine. Health Technol 2018 Jul 27;9(1):1-6. [doi: 10.1007/s12553-018-0244-4]
5. Choe J, Lee SM, Do K, Lee G, Lee J, Lee SM, et al. Deep learning-based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. Radiology 2019 Aug;292(2):365-373. [doi: 10.1148/radiol.2019181960] [Medline: 31210613]
6. Sivasamy J, Subashini T. Classification and predictions of lung diseases from chest x-rays using MobileNet. The International Journal of Analytical and Experimental Modal Analysis 2020 Mar;12(3):665-672 [FREE Full text]
7. Wang Y, Zhang Y, Xuan W, Kao E, Cao P, Tian B, et al. Fully automatic segmentation of 4D MRI for cardiac functional measurements. Med Phys 2019 Jan;46(1):180-189 [FREE Full text] [doi: 10.1002/mp.13245] [Medline: 30352129]
8. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. CheXNet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv. Preprint posted online on December 25, 2017 [FREE Full text]
9. Salehinejad H, Valaee S, Dowdell T, Colak E, Barfett J. Generalization of deep neural networks for chest pathology classification in X-rays using generative adversarial networks. : IEEE; 2018 Presented at: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); April 15-20, 2018; Calgary, AB, Canada p. 990-994. [doi: 10.1109/icassp.2018.8461430]
10. Chandra TB, Verma K. Pneumonia detection on chest x-ray using machine learning paradigm. In: Chaudhuri B, Nakagawa M, Khanna P, Kumar S, editors. roceedings of 3rd International Conference on Computer Vision and Image Processing. Advances in Intelligent Systems and Computing, vol 1022. Singapore: Springer; 2019:21-33.
11. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. : IEEE; 2017 Presented at: Proceedings of the IEEE conference on computer vision and pattern recognition; July 21-26, 2017; Honolulu, HI p. 2097-2106. [doi: 10.1109/cvpr.2017.369]
12. Smit A, Jain S, Rajpurkar P, Pareek A, Ng AY, Lungren MP. Combining automatic labelers and expert annotations for accurate radiology report labeling using BERT. In: ACL Anthology.: Association for Computational Linguistics; 2020 Presented at: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP); November 2020; Online p. 1500-1519. [doi: 10.18653/v1/2020.emnlp-main.117]
13. Ho TK, Gwak J. Multiple feature integration for classification of thoracic disease in chest radiography. Applied Sciences 2019 Oct 02;9(19):4130. [doi: 10.3390/app9194130]
14. Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 2020 Jun;43(2):635-640 [FREE Full text] [doi: 10.1007/s13246-020-00865-4] [Medline: 32524445]
15. Abbas A, Abdelsamea M, Gaber MM. Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Appl Intell 2020 Sep 05;51(2):854-864. [doi: 10.1007/s10489-020-01829-7]
16. Chen J, Wu L, Zhang J, Zhang L, Gong D, Zhao Y, et al. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. medRxiv. Preprint posted online on March 01, 2020 [FREE Full text] [doi: 10.1101/2020.02.25.20021568]
17. Narin A, Kaya C, Pamuk Z. Automatic detection of Coronavirus disease (covid-19) using X-ray images and deep convolutional neural networks. arXiv. Preprint posted online on March 24, 2020 [FREE Full text]

18. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. Radiology 2020 Aug;296(2):E65-E71 [FREE Full text] [doi: 10.1148/radiol.2020200905] [Medline: 32191588]
19. Sethy PK, Behera SK. Detection of Coronavirus disease (covid-19) based on deep features. Preprints. Preprint posted online on March 18, 2020 [FREE Full text] [doi: 10.20944/preprints202003.0300.v1]
20. Hemdan ED, Shouman MA, Karar ME. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in X-ray images. arXiv. Preprint posted online on March 24, 2020 [FREE Full text]
21. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, et al. A deep learning algorithm using ct images to screen for Coronavirus disease (covid-19). medRxiv. Preprint posted online on March 11, 2020 [FREE Full text]
22. Brunese L, Mercaldo F, Reginelli A, Santone A. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from x-rays. Comput Methods Programs Biomed 2020 Nov;196:105608 [FREE Full text] [doi: 10.1016/j.cmpb.2020.105608] [Medline: 32599338]
23. Wang L, Lin ZQ, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci Rep 2020 Nov 11;10(1):19549 [FREE Full text] [doi: 10.1038/s41598-020-76550-z] [Medline: 33177550]
24. Song Y, Zheng S, Li L, Zhang X, Ziwang Huang Z, Chen H, et al. Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. medRxiv. Preprint posted online on February 25, 2020 [FREE Full text] [doi: 10.1101/2020.02.23.20026930]
25. Zheng C, Deng X, Fu Q, Zhou Q, Feng J, Ma H, et al. Deep learning-based detection for COVID-19 from chest CT using weak label. medRxiv. Preprint posted online on March 26, 2020 [FREE Full text] [doi: 10.1101/2020.03.12.20027185]
26. Xu X, Jiang C, Ma C. Deep learning system to screen coronavirus disease 2019 pneumonia. arXiv. Preprint posted online on February 21, 2020 [FREE Full text]
27. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med 2020 Jun;121:103792 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.103792] [Medline: 32568675]
28. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Comput Biol Med 2020 Jun;121:103795 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.103795] [Medline: 32568676]
29. JSRT chest X-ray dataset. Japanese Society of Radiological Technology. URL: http://db.jsrt.or.jp/eng-04.php [accessed 2021-02-04]
30. SCR Segmentation Maps for JSRT Chest X-ray Dataset. SCR Segmentation Maps for JSRT Chest X-ray Dataset. SCR Database: Download. URL: http://www.isi.uu.nl/Research/Databases/SCR/download.php [accessed 2021-02-04]
31. Cohen JP, Morrison P, Dao L, Roth K, Duong TQ, Ghassemi M. COVID-19 image data collection: prospective predictions are the future. arXiv. Preprint posted online on June 22, 2020. https://github.com/ieee8023/covid-chestxray-dataset [FREE Full text]
32. NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories. National Institutes of Health - Clinical Center. URL: https://academictorrents.com/details/557481faacd824c83fbf57dcf7b6da9383b3235a [accessed 2021-02-04]
33. Zoph B, Vasudevan V, Shlens J, Le QV. Learning Transferable Architectures for Scalable Image Recognition. : IEEE; 2018 Presented at: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; June 18-23, 2018; Salt Lake City, UT p. 8697-8710 URL: https://ieeexplore.ieee.org/document/8579005 [doi: 10.1109/cvpr.2018.00907]
34. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. : IEEE; 2017 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); July 21-26, 2017; Honolulu, HI p. 1800-1807 URL: https://ieeexplore.ieee.org/document/8099678 [doi: 10.1109/cvpr.2017.195]
35. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. : IEEE; 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV p. 2818-2826 URL: https://ieeexplore.ieee.org/document/7780677 [doi: 10.1109/CVPR.2016.308]
36. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-ResNet and the impact of residual connections on learning. : AAAI Press; 2017 Presented at: AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence; February 2017; San Francisco, CA p. 4278-4284 URL: https://dl.acm.org/doi/10.5555/3298023.3298188
37. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. : IEEE; 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vega, NV p. 770-778 URL: https://ieeexplore.ieee.org/document/7780459 [doi: 10.1109/cvpr.2016.90]

Abbreviations

AUC: area under the curve
CNN: convoluted neural network
CT: computed tomography
CXR: chest x-ray

DCGAN: Deep Convolutional Generative Adversarial Network
NIH: National Institute of Health
SARS: severe acute respiratory syndrome

Edited by R Kukafka, C Basch; submitted 19.08.20; peer-reviewed by I Apostolopoulos, I Gabashvili; comments to author 24.09.20; revised version received 12.10.20; accepted 31.01.21; published 10.02.21.

Please cite as:
Albahli S, Yar GNAH
Fast and Accurate Detection of COVID-19 Along With 14 Other Chest Pathologies Using a Multi-Level Classification: Algorithm Development and Validation Study
J Med Internet Res 2021;23(2):e23693
URL: http://www.jmir.org/2021/2/e23693/
doi:10.2196/23693
PMID:33529154

Original Paper

A Machine Learning Prediction Model of Respiratory Failure Within 48 Hours of Patient Admission for COVID-19: Model Development and Validation

Siavash Bolourani[1], MD; Max Brenner[1], MD, PhD; Ping Wang[1], MD; Thomas McGinn[1], MD, MPH; Jamie S Hirsch[1], MD; Douglas Barnaby[1*], MD; Theodoros P Zanos[1*], PhD, MSc, BEng; Northwell COVID-19 Research Consortium[2*]

[1]Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, United States

[2]See Acknowledgments

[*]these authors contributed equally

Corresponding Author:
Theodoros P Zanos, PhD, MSc, BEng
Feinstein Institutes for Medical Research
Northwell Health
350 Community Dr
Room 1257
Manhasset, NY, 11030
United States
Phone: 1 5165620484
Email: tzanos@northwell.edu

Abstract

Background: Predicting early respiratory failure due to COVID-19 can help triage patients to higher levels of care, allocate scarce resources, and reduce morbidity and mortality by appropriately monitoring and treating the patients at greatest risk for deterioration. Given the complexity of COVID-19, machine learning approaches may support clinical decision making for patients with this disease.

Objective: Our objective is to derive a machine learning model that predicts respiratory failure within 48 hours of admission based on data from the emergency department.

Methods: Data were collected from patients with COVID-19 who were admitted to Northwell Health acute care hospitals and were discharged, died, or spent a minimum of 48 hours in the hospital between March 1 and May 11, 2020. Of 11,525 patients, 933 (8.1%) were placed on invasive mechanical ventilation within 48 hours of admission. Variables used by the models included clinical and laboratory data commonly collected in the emergency department. We trained and validated three predictive models (two based on XGBoost and one that used logistic regression) using cross-hospital validation. We compared model performance among all three models as well as an established early warning score (Modified Early Warning Score) using receiver operating characteristic curves, precision-recall curves, and other metrics.

Results: The XGBoost model had the highest mean accuracy (0.919; area under the curve=0.77), outperforming the other two models as well as the Modified Early Warning Score. Important predictor variables included the type of oxygen delivery used in the emergency department, patient age, Emergency Severity Index level, respiratory rate, serum lactate, and demographic characteristics.

Conclusions: The XGBoost model had high predictive accuracy, outperforming other early warning scores. The clinical plausibility and predictive ability of XGBoost suggest that the model could be used to predict 48-hour respiratory failure in admitted patients with COVID-19.

(J Med Internet Res 2021;23(2):e24246) doi:10.2196/24246

KEYWORDS

artificial intelligence; prognostic; model; pandemic; severe acute respiratory syndrome coronavirus 2; modeling; development; validation; COVID-19; machine learning

Introduction

On March 11, 2020, COVID-19, the disease caused by SARS-CoV-2 infection, was declared a pandemic by the World Health Organization [1]. As of December 16, 2020, there were more than 17 million confirmed COVID-19 cases and over 300,000 deaths in the United States. During the first wave, New York was the epicenter of the pandemic in the United States, with over 390,000 cases and 30,000 deaths before the summer [2].

Respiratory failure is the leading cause of death among patients with COVID-19, with up to one-third of patients admitted with COVID-19 requiring invasive mechanical ventilation (IMV) [3-8]. The decision to initiate IMV in these patients is not straightforward. Many patients with severe disease appear comfortable despite profound hypoxemia, and they are commonly managed with supplemental oxygen, self-proning, and close monitoring [9,10]. However, some of these patients subsequently deteriorate and require IMV following transfer from the emergency department (ED). This subgroup has worse outcomes than those placed on IMV initially [11]. Before the surge of COVID-19, patients initially admitted to a non–critical care setting who needed an unplanned transfer to an intensive care unit (ICU) had greater morbidity and mortality than those admitted directly to a critical care unit [12-14]. Thus, accurately identifying patients at high risk for deterioration could improve clinical outcomes as a result of closer monitoring, direct admission to a critical care unit, or earlier discussions regarding patient preferences and goals of care.

Methods of identifying patients at high risk for or in the early stages of clinical deterioration have been actively researched for decades. The field has generated many severity-of-illness calculators, early warning scores, and, more recently, predictive analytic tools that use advanced machine learning and artificial intelligence [15-23]. Our goal was to derive a prediction model that estimates the risk of short-term (<48 hours) respiratory failure for patients with COVID-19 who were not initially placed on IMV. Such a tool could improve outcomes by avoiding delayed admission to a critical care unit, resulting in the provision of additional respiratory support and closer monitoring, or the initiation of earlier discussions around the goals of care.

Methods

Overview

This retrospective observational cohort drew data from 13 acute care hospitals of Northwell Health, the largest health care system in New York State. Data were extracted from the electronic health record (EHR) Sunrise Clinical Manager (Allscripts). EHRs were screened for adult patients (aged ≥21 years) who received a positive test result for SARS-CoV-2 based on a nasopharyngeal sample tested using polymerase chain reaction assays. Included patients were hospitalized and were discharged, died, or spent a minimum of 48 hours in the hospital between March 1, 2020, and May 11, 2020. For patients who had multiple qualifying hospital admissions, only the first hospitalization was included. Patients who were transferred between hospitals within the health system were treated as one hospital encounter. A total of 11,919 patients were identified. Patients were excluded if they were placed on mechanical ventilation prior to inpatient admission. A total of 11,525 patients remained for analysis. The Institutional Review Board of Northwell Health approved the study protocol and waived the requirement for informed consent.

Data Acquisition

Data collected from EHRs included patient demographics, comorbidities, home medications, initial vitals and laboratory values, treatments (eg, oxygen therapy, mechanical ventilation), and clinical outcomes (eg, length of stay, discharge, mortality). Vitals and laboratory testing were restricted to those obtained while the patient was in the ED.

Outcomes

The target outcome variable was defined as intubation and mechanical ventilation within 48 hours of admission. In the EHR, the admission time was recorded, and the intubation event was defined as the first time mechanical ventilation was recorded.

Predictive Machine Learning Model

We evaluated three predictive models: XGBoost, XGBoost + SMOTEENN (combined oversampling using SMOTE and undersampling using edited nearest neighbors) [24], and logistic regression [25]. XGBoost combines a recursive gradient–boosting method, called Newton boosting, with a decision-tree model. Given that each tree is boosted in parallel, the model efficiently provides accurate predictions [26]. Furthermore, because each tree is boosted recursively and in parallel, the model benefits from the high interpretability of the variable importance features.

The XGBoost + SMOTEENN method involves combined oversampling using SMOTE and undersampling using edited nearest neighbors on the training set before training an XGBoost model [27]. This method has been shown to have the best performance in the resampling data sets [28]. Furthermore, in our experience, when using any of the oversampling or undersampling methods alone, calibration of the model is severely affected. However, when we combine oversampling the minority class with undersampling of the majority class, we found that we get a more accurate model both in terms of discriminability and minimizing the effect on the calibration of the model.

For every learning framework, we validated the model with external validation using each hospital (ie, for each fold, one hospital was picked as a testing set and the others as a training set). Only hospitals with >1000 patients with COVID-19 in the data set were picked for the testing sets, and a random sample of 1000 patients was picked to be our testing set for each fold. Grid search was used to hypertune the parameters of the respective models. The XGBoost model was tuned based on min_child_weight, gamma, subsample, colsample_bytree, and max_depth parameters, and the ranges of the values were 1-20, 0.5-20, 0.2-1.0, 0.2-1.0, and 2-40, respectively.

When data were missing, we imputed weighted k-nearest neighbors [29] for numerical values and added a category "missing" for categorical values. We used one-hot [30] to encode categorical variables as a one-hot numeric array. The most important variables were calculated based on a decrease in the mean Gini coefficient (ie, the variables most useful in splitting the data to help make a prediction) for XGBoost and XGBoost + SMOTEENN; and by the absolute value of the regression coefficient for logistic regression, and were calculated based on the largest hospital testing set. The resulting receiver operating characteristic (ROC) curves and corresponding accuracy, recall (sensitivity), specificity, geometric mean, and F_β-score were averaged. For the F_β-score, the β parameter value was designated as $\beta=4$ to capture a higher detriment of false negatives than false positives (ie, if we value recall, β times as much as the precision). For definitions of these measures and how they were calculated, see Multimedia Appendix 1.

Calibration curves (reliability curves) were plotted by dividing the testing sets (for each hospital fold) into 10 bins randomly with an increasing fraction of patients that had respiratory failure in the sample. The fraction positives (patients who had respiratory failure) and their mean corresponding predicted value from the corresponding model were depicted and averaged into 10 bins. The Brier score was calculated for each external hospital fold and the mean Brier score and standard deviation were calculated and depicted in the legend of the calibration curve. For further explanation of these measures and how they were calculated, see Multimedia Appendix 1.

Python 2.6 (Python Software Foundation) was used to implement our machine learning framework. The respective prediction models of XGBoost and logistic regression from the scikit-learn application programming interface (API) in Python were used [31]. GridSearchCV from the scikit-learn API was used to perform the grid search and hypertune the parameters. We used the default imblearn API version of the SMOTEENN [27]. SimpleImputer [32] was used for imputations, which were replaced with a new category, "missing." KNNImputer [33] was used to impute the missing numerical data [29]. The default value for k=5 was not changed. OneHotEncoder from the sklearn API was used to transform categorical variables to one-hot numeric arrays.

Modified Early Warning Score

The Modified Early Warning Score (MEWS) was computed from patient vital signs (Multimedia Appendix 2) and is a variant of other known and used risk scores [34,35]. The MEWS ranges from 0 to 15 and incorporates heart rate (beats per minute), respiratory rate (breaths per minute), systolic blood pressure (mm Hg), and body temperature (degrees Celsius). In our data set, one MEWS subcomponent, the AVPU (alert, verbal, pain, unresponsive) neurologic assessment, had a significant amount of missing data (>80%; data not shown) and was not included in the MEWS calculation for this project. An elevated MEWS indicates a risk for clinical instability, including death or the need for ICU admission [36]. In 2012, our health system created a custom modification that was incorporated into the EHR. It includes automatic calculation and display of MEWS and other modules via Arden Syntax Medical Logic Modules [37]. Based on local health system guidelines, any score ≥7 requires an escalation in intensity of care. For example, MEWS >7 requires increased frequency of vital sign measurement (every 2 hours), MEWS >8 requires evaluation by a licensed independent provider, MEWS >9 requires consideration of evaluation by a rapid response team, and MEWS >10 requires a change in the level of service per a defined protocol. For the MEWS, we chose the highest value the patient had while in the ED.

Results

Patient Characteristics

During the study period, we identified 11,525 patients admitted from the ED with a diagnosis of COVID-19. Of these, 933 (8.0%) were placed on IMV within 48 hours of admission. Baseline characteristics (demographics, baseline vital signs, and laboratory measurements) for all patients are shown in Table 1, stratified by study outcome. Comorbidities were captured from ICD-10 codes listed in the EHR.

Table 1. Demographic, clinical, and laboratory data from hospitalized patients.

Variables	Not intubated (n=10,592)	Intubated (n=933)	Missing (%)
Demographic characteristics			
Age (years), median (IQR)	65.00 (54.00-77.00)	66.00 (56.00-75.00)	0
Female, n (%)	4530 (42.8)	327 (35.0)	0
Primary language, English, n (%)	8498 (80.2)	746 (80.0)	0
Race, n (%)			0
Black	2199 (20.8)	236 (25.3)	N/A[a]
Asian	889 (8.4)	77 (8.3)	N/A
White	4148 (39.2)	310 (33.2)	N/A
Declined	71 (0.7)	8 (0.9)	N/A
Other	2884 (27.2)	268 (28.7)	N/A
Unknown	401 (3.8)	34 (3.6)	N/A
Ethnicity, n (%)			0.1
Hispanic or Latino	2238 (21.1)	202 (21.7)	N/A
Not Hispanic or Latino	7685 (72.6)	648 (69.5)	N/A
Declined	43 (0.4)	1 (0.1)	N/A
Unknown	618 (5.8)	82 (8.8)	N/A
Vital signs			
Systolic blood pressure (mm Hg), median (IQR)	134.00 (118.00-150.00)	134.00 (115.00-151.75)	0.5
Diastolic blood pressure (mm Hg), median (IQR)	79.00 (70.50-87.00)	77.00 (69.00-86.00)	0.6
Heart rate (beats/minute), median (IQR)	94.00 (85.00-102.00)	97.00 (88.50-112.00)	0.4
Respiratory rate (breaths/minute), median (IQR)	21.00 (18.00-25.00)	24.00 (20.00-32.00)	0.8
Temperature (°C), mean (SD)	37.77 (0.97)	37.86 (1.11)	1.6
Oxygen saturation (%), median (IQR)	97.00 (95.00-98.00)	96.00 (93.00-98.00)	1.7
BMI, mean (SD)	29.12 (7.79)	30.39 (9.21)	47.1
Laboratory data			
White blood cell count ($\times 10^9$/L), median (IQR)	7.34 (5.45-9.92)	8.25 (6.20-11.50)	9
Absolute neutrophil count ($\times 10^9$/L), median (IQR)	5.68 (3.95-8.11)	6.84 (4.76-9.62)	11.5
Absolute lymphocyte count ($\times 10^9$/L), median (IQR)	0.90 (0.63-1.27)	0.80 (0.56-1.13)	11.5
Hemoglobin (g/dL), mean (SD)	12.93 (2.12)	13.14 (2.11)	9
Platelets (K/µL), mean (SD)	230.17 (101.93)	217.19 (87.45)	9.1
Sodium (mmol/L), mean (SD),	136.64 (6.21)	135.38 (5.74)	11.9
Carbon dioxide (mmol/L), mean (SD)	23.61 (3.79)	22.67 (4.68)	11.9
Creatinine (mg/dL), median (IQR)	1.03 (0.80-1.46)	1.20 (0.92-1.75)	12
Bilirubin (mg/dL), median (IQR)	0.50 (0.40-0.70)	0.60 (0.40-0.80)	12.5
Ferritin (ng/mL), mean (SD)	1283.50 (2732.65)	1731.05 (2631.38)	73.2
Procalcitonin (ng/mL), mean (SD)	1.22 (10.96)	2.12 (8.16)	66.3
D-dimer (ng/mL), mean (SD)	1871.84 (5306.42)	2659.09 (6798.96)	65.4
Lactate dehydrogenase (U/L), mean (SD)	455.61 (213.04)	611.05 (272.16)	71
pH (arterial), mean (SD)	7.42 (0.09)	7.39 (0.11)	96.7
Partial pressure of oxygen (arterial, mm Hg), mean (SD)	99.90 (65.17)	85.26 (61.42)	94.8

Variables	Not intubated (n=10,592)	Intubated (n=933)	Missing (%)
Partial pressure of carbon dioxide (arterial, mm Hg), mean (SD)	34.66 (9.38)	35.38 (11.45)	94.7
Comorbidities			
Hypertension, n (%)	1183 (11.2)	115 (12.3)	0
Diabetes, n (%)	685 (6.5)	77 (8.3)	0
Coronary artery disease, n (%)	148 (1.4)	15 (1.6)	0
Asthma/chronic obstructive pulmonary disease, n (%)	242 (2.3)	20 (2.1)	0
Chronic kidney disease, n (%)	99 (0.9)	8 (0.9)	0
HIV, n (%)	26 (0.2)	1 (0.1)	0

[a]N/A: not applicable.

Prediction Models for Respiratory Failure

Based on XGBoost, the mean area under the curve (AUC) of the ROC (AUCROC) curve was 0.77 (SD 0.05) and the mean AUC of the PR curve (AUCPR) was 0.26 (SD 0.04; Figure 1). The 10 most important variables, in order of decreasing importance, were as follows: most invasive mode of oxygen delivery being a nonrebreather mask, Emergency Severity Index (ESI) values of 1 and 3, maximum respiratory rate, maximum oxygen saturation, Black race, age on admission, eosinophil percentage, serum sodium level, and serum lactate level (Figure 1). The confusion matrix for the model's largest hospital testing set showed that most false predictions were false negatives (those who were predicted to not require intubation but were intubated within 48 hours). False positives (those who were predicted to require intubation but were not intubated within 48 hours) were the minority of predictions (Figure 1). The model had a mean accuracy of 0.919 (SD 0.028). The corresponding mean precision, recall, specificity, geometric mean, and F_β-score were 0.521 (SD 0.329), 0.051 (SD 0.030), 0.994 (SD 0.005), 0.337 (SD 0.042), and 0.054 (SD 0.029), respectively (Table 2).

Based on the XGBoost + SMOTEENN model, the mean AUCs of the ROC and PR curves were 0.76 (SD 0.03) and 0.24 (SD 0.06), respectively (Figure 2). The 10 most important variables, in order of decreasing importance, were as follows: most invasive mode of oxygen delivery being a nonrebreather mask, ESI value of 3, male gender, White race, minimum respiratory rate, Black race, ESI value of 2, most invasive mode of oxygen delivery being nasal cannula, ESI value of 1, and Hispanic ethnicity (Figure 2). The mean confusion matrix showed that most false predictions were false positives (those who were predicted to require intubation but were not intubated within 48 hours). False negatives (those who were predicted to not require intubation but were intubated within 48 hours) were the minority of predictions (Figure 2). Although this model did not have the highest accuracy, it achieved the highest mean recall, geometric mean, and F_β-score of 0.228 (SD 0.095), 0.508 (SD 0.063), and 0.226 (SD 0.010), respectively. The corresponding mean accuracy, precision, and specificity were 0.893 (SD 0.016), 0.303 (SD 0.089), and 0.955 (SD 0.005), respectively (Table 2).

Figure 1. The XGBoost model for predicting respiratory failure within 48 hours. (A) ROC curve and (B) PR curve based on a cross-hospital validation performed by leaving a hospital out as a testing set and using the rest in the training set. Only hospitals with >1000 patients with COVID-19 were selected for testing sets. The mean ROC and PR curves are shown in dark blue and their corresponding standard deviations are shown in gray. The MEWS metrics are shown in light yellow. (C) Measurement of the 10 variables with the highest relative importance based on the amount they reduced the Gini coefficient for the largest hospital testing set. (D) Confusion matrix visually represents the predicted values versus actual prediction for the largest hospital testing set. AUC: area under the curve of ROC; AUCPR: area under the curve of the precision-recall curve; ED: emergency department; LIJ: Long Island Jewish; MEWS: Modified Early Warning Score; PR: precision-recall; ROC: receiver operating characteristic.

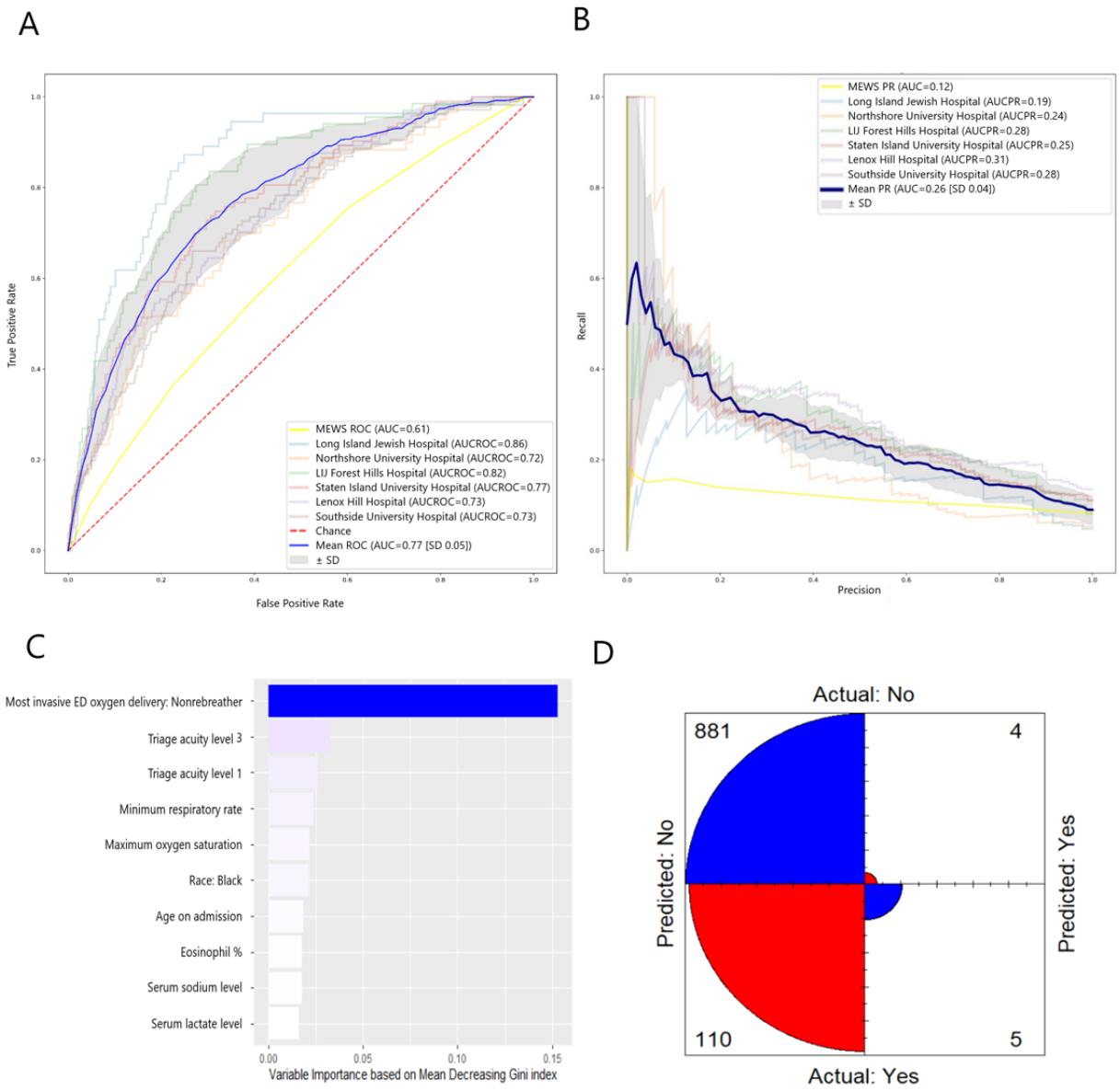
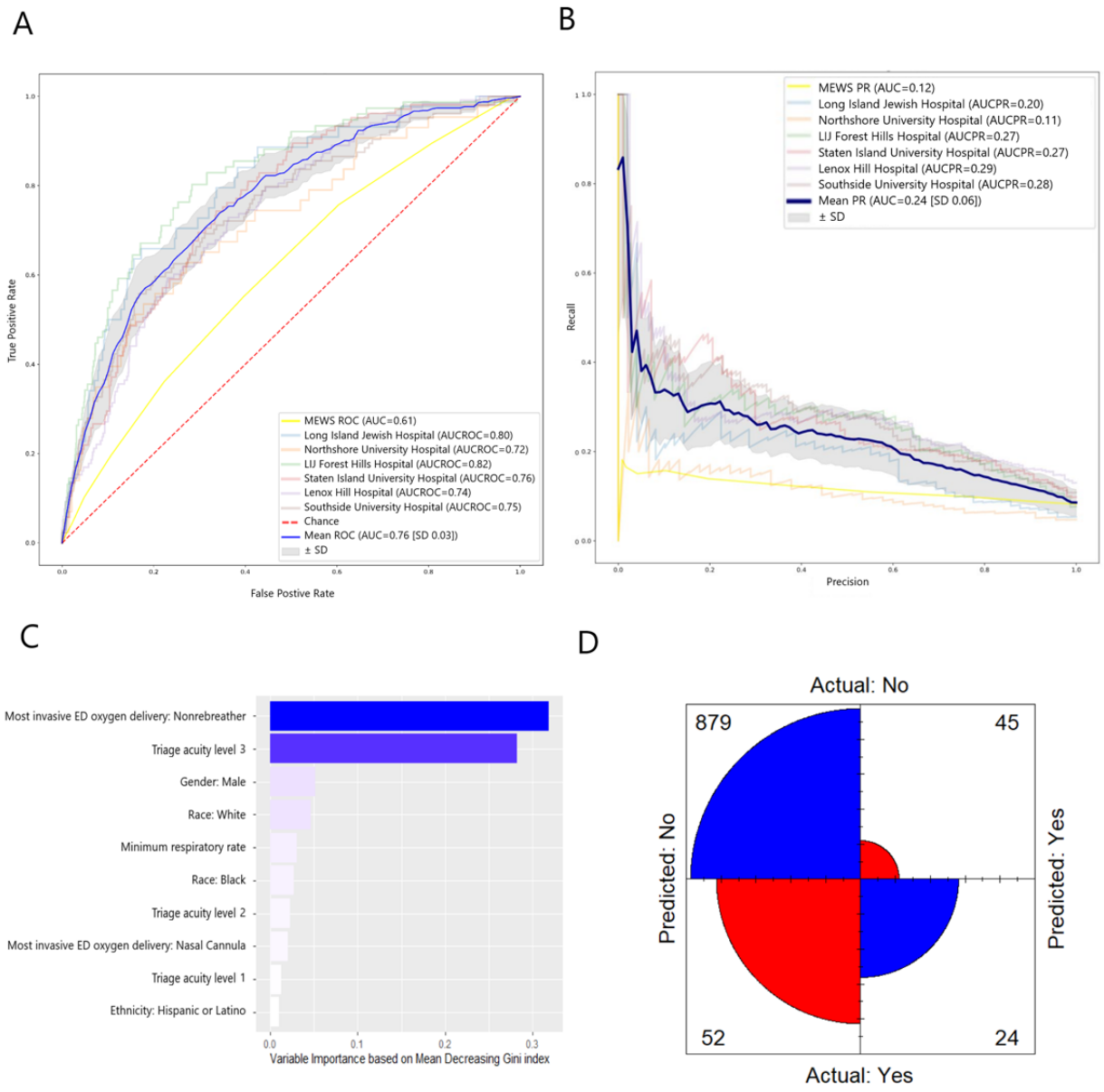


Table 2. Mean area under the curve of the receiver operating characteristic curve, area under the curve of the precision-recall curve, accuracies, precisions, recalls, specificities, geometric means, and Fβ-score (β=4) for models investigated.

Measure	XGBoost, mean (SD)	XGBoost + SMOTEENN, mean (SD)	Logistic regression, mean (SD)	Modified Early Warning Score
Area under the curve of the receiver operating characteristic curve	0.77 (0.05)	0.76 (0.03)	0.70 (0.05)	0.61
Area under the curve of the precision-recall curve	0.26 (0.04)	0.24 (0.06)	0.18 (0.06)	0.12
Accuracy	0.919 (0.028)	0.893 (0.016)	0.915 (0.027)	0.913
Precision	0.521 (0.329)	0.303 (0.089)	0.322 (0.375)	0.165
Recall	0.051 (0.030)	0.228 (0.095)	0.009 (0.013)	0.017
Specificity	0.994 (0.005)	0.955 (0.005)	0.998 (0.002)	0.992
Geometric mean	0.337 (0.042)	0.506 (0.063)	0.285 (0.051)	0.296
F_β-score	0.054 (0.029)	0.226 (0.088)	0.010(0.014)	0.018

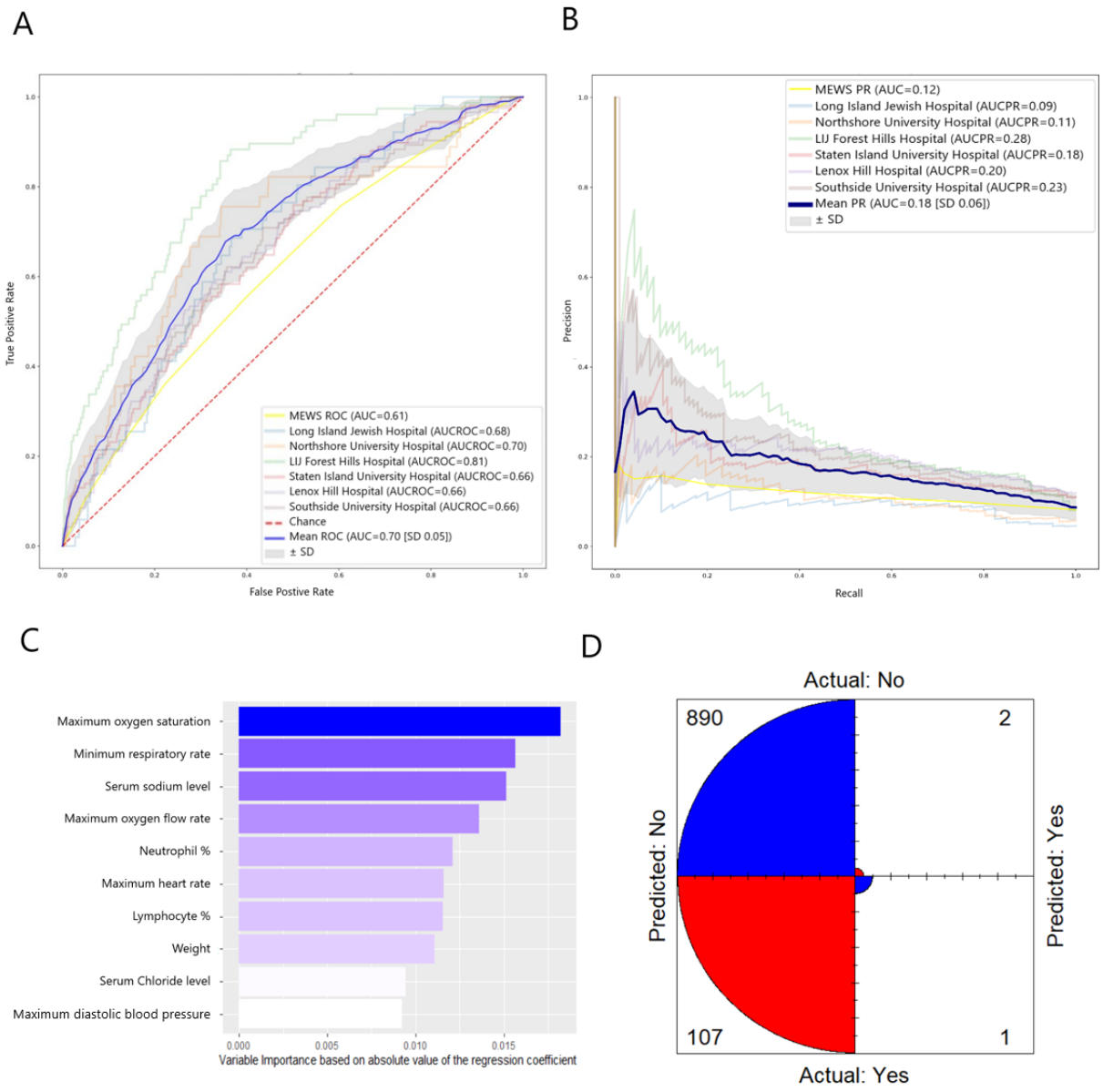
Figure 2. The XGBoost + SMOTEENN model for predicting respiratory failure within 48 hours. (A) ROC curve and (B) PR curve based on a cross-hospital validation performed by leaving one hospital out as a testing set and using the remaining hospitals for the training set. Only hospitals with >1000 patients with COVID-19 were selected for testing sets. The mean ROC and PR curves are shown in dark blue and their corresponding standard deviations are shown in gray. The MEWS metrics are shown in light yellow. (C) The 10 variables with the highest relative importance measured by the amount the variable reduced the Gini coefficient. (D) Mean confusion matrix visually represents the predicted values versus actual prediction. AUC: area under the curve of ROC; AUCPR: area under the curve of the precision-recall curve; ED: emergency department; LIJ: Long Island Jewish; MEWS: Modified Early Warning Score; PR: precision-recall; ROC: receiver operating characteristic.



We also examined the performance of a logistic regression model. The mean AUCs of the ROC and PR curves were 0.70 (SD 0.05) and 0.18 (SD 0.06), respectively. Mean accuracy, precision, recall, specificity, geometric mean, and F_β-score were 0.915 (SD 0.027), 0.322 (SD 0.375), 0.009 (SD 0.013), 0.994 (SD 0.005), 0.285 (SD 0.051), and 0.010 (SD 0.014), respectively (Figure 3 and Table 2). MEWS was used to compare ROC and PR curves. MEWS resulted in AUCs of the ROC and PR curves of 0.61 and 0.12, respectively (Figures 1-3). For MEWS, accuracy, precision, recall, specificity, geometric mean, and F_β-score were 0.913, 0.165, 0.017, 0.992, 0.296, and 0.018, respectively.

The calibration curves showed that all three models were well calibrated among all hospital folds, although all three deviated from perfect calibration as the fraction of positives increased (Figure 3). The corresponding mean Brier score for XGBoost, XGBoost + SMOTEENN, and logistic regression was 0.071 (SD 0.019), 0.079 (SD 0.016), and 0.077 (SD 0.018), respectively (Figure 3).

Figure 3. The logistic regression model for predicting respiratory failure within 48 hours. (A) ROC curve and (B) PR curve based on a cross-hospital validation performed by leaving a hospital out as a testing set and using the rest for the training set. Only hospitals with >1000 patients with COVID-19 were selected for testing sets. The mean ROC and PR curves are shown in dark blue and their corresponding standard deviations are shown in gray. The MEWS metrics are shown in light yellow. (C) The 10 variables with the highest relative importance measured by the absolute value of the regression coefficient. (D) Mean confusion matrix visually represents the predicted values versus actual prediction. AUC: area under the curve of ROC; AUCPR: area under the curve of the precision-recall curve; LIJ: Long Island Jewish; MEWS: Modified Early Warning Score; PR: precision-recall; ROC: receiver operating characteristic.



Discussion

We presented three models (two of which were based on XGBoost) for predicting early respiratory failure in patients given a diagnosis of COVID-19 and admitted to the hospital from the ED. One model was tilted toward precision and specificity (XGBoost) and the other was tilted toward recall (XGBoost + SMOTEENN). These models are based on baseline characteristics, ED vital signs, and laboratory measurements. Using an automated tool to estimate the probability of respiratory failure could identify at-risk patients for earlier interventions (eg, closer monitoring, critical care consultation, earlier discussions about goals of care) and improve patient outcomes.

We evaluated three machine learning models: XGBoost, XGBoost + SMOTEENN, and logistic regression [38-40]. XGBoost is widely used due to its high efficiency and predictability, and it has been used to predict health care outcomes in patients with [41,42] and without [43-45] COVID-19. In our study, XGBoost was the most accurate prediction model, with an accuracy of 0.919 (SD 0.028) and precision of 0.521 (SD 0.329; Figure 1), similar to the findings of another study that examined combined outcomes [46]. However, what is different in our model is that it achieves

cross-hospital validation. Such accuracy showcases the ability of the model to separate intubations from nonintubations within the 48-hour window of interest. Such a model would be useful for physicians as it more accurately and consistently identifies patients at high risk for intubation.

We also constructed an XGBoost + SMOTEENN model. SMOTEENN was used to improve the sensitivity of our prediction, as our data set was imbalanced (ie, only ~8% of our COVID-19 cohort were intubated), while keeping deviation from accuracy and calibration of the model to a minimum. Compared to XGBoost, the XGBoost + SMOTEENN model had lower accuracy and precision, but greater recall (or sensitivity; 0.228 [SD 0.095]; Figure 2). This higher sensitivity can identify more patients who require IMV, suggesting that this model may be more suitable for broad or automated screening of patients.

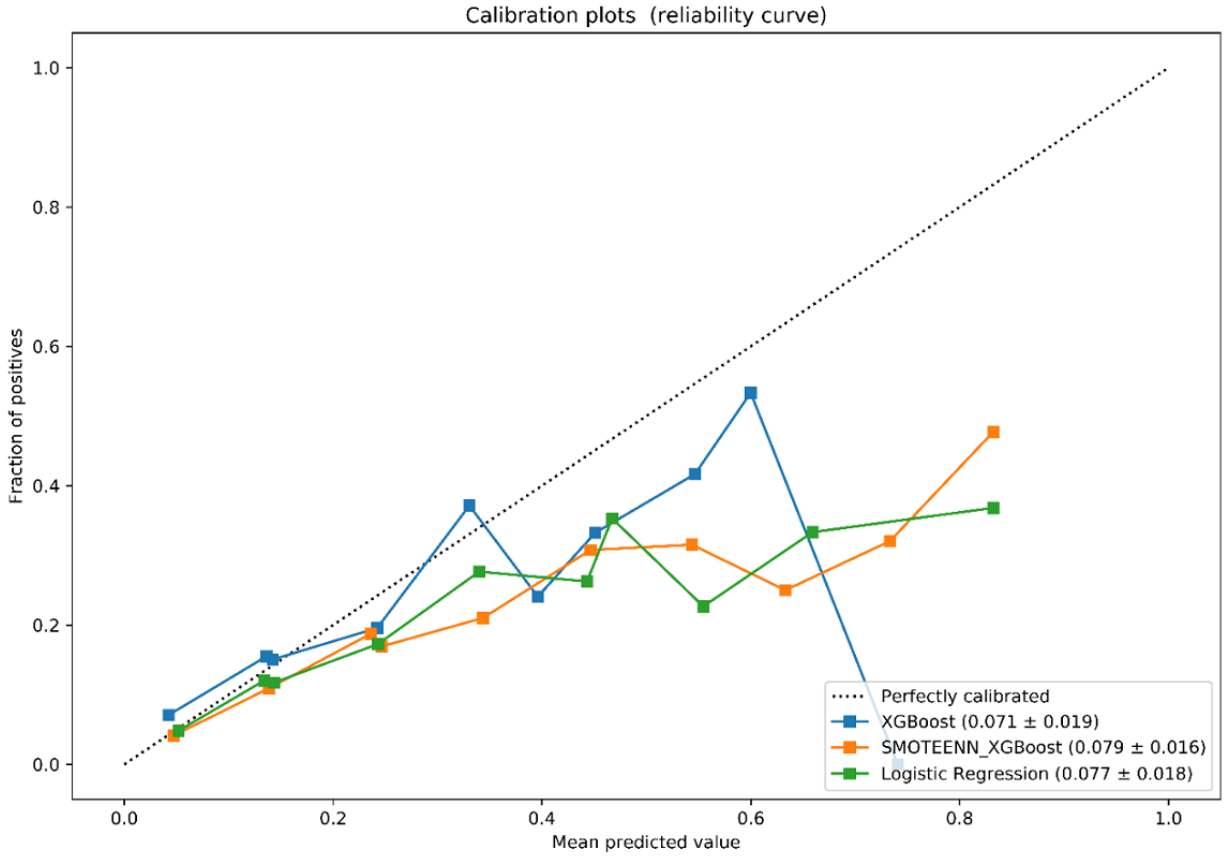
We also examined the performance of a logistic regression model to determine whether a compact, linear model could accurately predict patient risk (Figure 3 and Table 2). Model performance was inferior to the XGBoost model. This supports earlier reports that machine learning techniques outperform classic models of logistic regression in their ability to predict many prognostic and health outcomes [47-49]. Finally, we compared the performance of our predictive machine learning models to the widely used MEWS [36]. MEWS was inferior to all three models described above in most of the measures examined.

Using the most important variables for our models, we identified clinically relevant measures that can best inform clinical decision making (Figures 1, 2). The XGBoost model was accurate and precise, as reflected by the low number of false positives of the model predictions (Figure 1). A more sensitive alternative to this model would be the XGBoost + SMOTEENN model, which had fewer false negatives than XGBoost (Figure 2). Both models share important predictors, such as information about the mode of oxygen delivery, triage acuity, demographic information, and respiratory rate. However, XGBoost (the more accurate model with higher precision) adds serum lactate, sodium, and eosinophil percentage to the top 10 most important variables. This indicates that when precision is important, measures such as lactate can rule out the most severe cases by becoming strong predictors. Among hospitals in Northwell Health, certain hospitals such as Long Island Jewish (which is one of the largest in terms of number of patients with COVID-19) had a high drop in their predictive ability when logistic regression was used. When Long Island Jewish was being validated, the 0.86 AUCROC of the XGBoost model dropped to 0.68 for logistic regression. This could partially be due to the nature of the outcome predicted (choice of ventilation from hospital staff), where one would expect different hospitals to possibly exhibit higher variability, not only for patient demographics, but also for hospital staff therapy choices.

Variable importance metrics revealed that the linear logistic regression models use laboratory variables primarily, whereas nonlinear XGBoost-based models prioritize clinical and demographic variables that better capture hospital-specific behavior (eg, oxygen delivery types prior to intubation) and increase the robustness of the model. However, we need to validate whether providing these variables along with the probability of respiratory failure would decrease the rate of identifying at-risk patients. Further prospective studies and randomized clinical trials are needed for this validation.

When examining the calibration of the models (Figure 4), we found that all models were well calibrated, yet as the fraction of positive cases increased, calibration suffered. This suggests that if a specific population of patients has a greater likelihood of intubation (eg, those aged >70 years, or with specific comorbid conditions), the model would need to be retrained to increase its accuracy and calibration.

Figure 4. Calibration plots (reliability curve) of the XGBoost, XGBoost + SMOTEENN, and logistic regression models for respiratory failure within 48 hours. Calibration is based on the precision probability (using predict_proba in Python). For creating the plots, sklearn.calibration.CalibratedClassifierCV (in Python) was used by inserting a fraction of positives and mean predicted values into 10 bins with an increasing fraction of positives (respiratory failures) for each hospital fold. The mean Brier score (SD) across all hospitals tested corresponding to the model is shown in the figure legend in parentheses.



Our study has several limitations. We extracted data on intubation timing from our EHR, which may have minor inaccuracies. Although a consistent temporal inaccuracy could create bias in underestimating/overestimating the intubation rate, we believe that these small inaccuracies are overcome by the average calculated from our large number of cases. Another limitation is that we relied on data from a multicenter, single health system for both implementation and validation. Thus, we were unable to externally validate the models in other health systems and hospitals with different protocols, which might affect the model's performance. In addition, because the study is retrospective, we can only suggest associations and correlations rather than identify the main contributors that lead to intubation and mechanical ventilation. Furthermore, the numerical missing variables were imputed with weighted k-nearest neighbors. Thus, the conclusions made from these variables assume uniformity in patient data based on those missing values. In the case of nonuniformity, the order of variable importance might change. Additionally, some clinical variables included in the model may appear to be obvious correlates of the clinical decision for intubation within 48 hours (eg, having nonrebreather oxygen flow as the most invasive form of ventilation). However, the association of all included variables is not deterministic: only 453 of 2633 patients on nonrebreather oxygen flow in the ED were intubated within 48 hours. In addition, given that these variables are available to clinicians and part of their decision making, we included them in our model. Finally, we used supervised learning on a homogenous database. Although we used cross-hospital validation and retrospectively validated our learning method, external generalizability of these learning methods to other health systems requires validation in prospective studies and randomized trials. Such high-quality evidence could provide more clues on clinical and economic impacts, as well as measures to improve them.

COVID-19 has evolved into an extremely challenging clinical and public emergency worldwide, especially in the New York City metropolitan area. As public health measures attempt to mitigate this disaster by slowing the spread and alleviating the heavy burden placed on health care systems, clinicians must make important decisions quickly and hospital administrators must manage resources and personnel. Furthermore, as predicted by many models [50-52], we are in the midst of a second wave of infection. Our models could inform clinical care by offering complementary performance characteristics (one model with superior recall, the other with greater precision) and supporting clinical decision making as we tackle this unprecedented public health crisis.

Acknowledgments

This work was supported by R24AG064191 from the National Institute on Aging, R01LM012836 from the National Library of Medicine, and R35GM118337 from the National Institute for General Medical Sciences (all National Institutes of Health).

Members of the Northwell COVID-19 Research Consortium include: Matthew Barish, Stuart L Cohen, Kevin Coppa, Karina W Davidson, Shubham Debnath, Lawrence Lau, Todd J Levy, Alexander Makhnevich, Marc D Paradis, and Viktor Tóth.

We acknowledge and honor all of our Northwell team members who consistently put themselves in harm's way during the COVID-19 pandemic. We dedicate this article to them, as well as all patients, as their vital contributions to knowledge about COVID-19 made it possible.

Authors' Contributions

SB, DPB, and TPZ conceptualized and designed the study. SB, JSH, and TPZ had full access to all data in the study and are responsible for the integrity of the data. SB and JSH performed data extraction and cleaning. MB, PW, TM, and JSH contributed to many discussions during manuscript development. SB, DPB, and TPZ contributed to drafts of the manuscript. SB trained and validated the models. SB and TPZ designed and created the figures. DBP and TPZ critically reviewed the paper, and PW and TPZ obtained funding. The Northwell COVID-19 Research Consortium prioritized this manuscript, organized meetings between contributing authors, and provided support in finalizing the manuscript for submission. All named authors read and approved the final submitted manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definitions of accuracy, precision, recall, specificity, geometric means, and Fβ-score.
[DOCX File , 8 KB - jmir_v23i2e24246_app1.docx]

Multimedia Appendix 2
Modified Early Warning Score calculation based on vital sign measurements.
[DOCX File , 8 KB - jmir_v23i2e24246_app2.docx]

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2021-02-01]
2. Coronavirus in the US: Latest Map and Case Count. The New York Times. 2020 Mar 3. URL: https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html [accessed 2020-05-08]
3. Richardson S, Hirsch J, Narasimhan M, Crawford J, McGinn T, Davidson K, the Northwell COVID-19 Research Consortium, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA 2020 May 26;323(20):2052-2059 [FREE Full text] [doi: 10.1001/jama.2020.6775] [Medline: 32320003]
4. Myers L, Parodi S, Escobar G, Liu V. Characteristics of Hospitalized Adults With COVID-19 in an Integrated Health Care System in California. JAMA 2020 Jun 02;323(21):2195-2198 [FREE Full text] [doi: 10.1001/jama.2020.7202] [Medline: 32329797]
5. Tsertsvadze T, Ezugbaia M, Endeladze M, Ratiani L, Javakhishvili N, Mumladze L, et al. Characteristics and outcomes of hospitalized adult COVID-19 patients in Georgia. MedRxiv Preprint published online on October 27, 2020. [doi: 10.1101/2020.10.23.20218255]
6. Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med 2020 May;46(5):846-848 [FREE Full text] [doi: 10.1007/s00134-020-05991-x] [Medline: 32125452]
7. Ruan Q, Yang K, Wang W, Jiang L, Song J. Correction to: Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med 2020 Jun;46(6):1294-1297 [FREE Full text] [doi: 10.1007/s00134-020-06028-z] [Medline: 32253449]
8. Goyal P, Choi J, Pinheiro L, Schenck E, Chen R, Jabri A, et al. Clinical Characteristics of Covid-19 in New York City. N Engl J Med 2020 Jun 11;382(24):2372-2374 [FREE Full text] [doi: 10.1056/NEJMc2010419] [Medline: 32302078]
9. Levitan R. Opinion: The Infection That's Silently Killing Coronavirus Patients. The New York Times. 2020 Apr 20. URL: https://www.nytimes.com/2020/04/20/opinion/sunday/coronavirus-testing-pneumonia.html [accessed 2020-05-08]

10. Tobin MJ. Basing Respiratory Management of COVID-19 on Physiological Principles. Am J Respir Crit Care Med 2020 Jun 01;201(11):1319-1320 [FREE Full text] [doi: 10.1164/rccm.202004-1076ED] [Medline: 32281885]
11. Kangelaris KN, Ware LB, Wang CY, Janz DR, Zhuo H, Matthay MA, et al. Timing of Intubation and Clinical Outcomes in Adults With Acute Respiratory Distress Syndrome. Crit Care Med 2016 Jan;44(1):120-129 [FREE Full text] [doi: 10.1097/CCM.0000000000001359] [Medline: 26474112]
12. Serin SO, Karaoren G, Esquinas AM. Delayed admission to ICU in acute respiratory failure: Critical time for critical conditions. Am J Emerg Med 2017 Oct;35(10):1571-1572. [doi: 10.1016/j.ajem.2017.04.026] [Medline: 28502761]
13. Renaud B, Santin A, Coma E, Camus N, Van Pelt D, Hayon J, et al. Association between timing of intensive care unit admission and outcomes for emergency department patients with community-acquired pneumonia. Crit Care Med 2009 Nov;37(11):2867-2874. [doi: 10.1097/CCM.0b013e3181b02dbb] [Medline: 19770748]
14. Liu V, Kipnis P, Rizk NW, Escobar GJ. Adverse outcomes associated with delayed intensive care unit transfers in an integrated healthcare system. J Hosp Med 2012 Mar;7(3):224-230. [doi: 10.1002/jhm.964] [Medline: 22038879]
15. Churpek MM, Carey KA, Dela Merced N, Prister J, Brofman J, Edelson DP. Validation of Early Warning Scores at Two Long-Term Acute Care Hospitals. Crit Care Med 2019 Dec;47(12):e962-e965 [FREE Full text] [doi: 10.1097/CCM.0000000000004026] [Medline: 31567342]
16. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. Crit Care Med 2016 Feb;44(2):368-374 [FREE Full text] [doi: 10.1097/CCM.0000000000001571] [Medline: 26771782]
17. Dziadzko MA, Novotny PJ, Sloan J, Gajic O, Herasevich V, Mirhaji P, et al. Multicenter derivation and validation of an early warning score for acute respiratory failure or death in the hospital. Crit Care 2018 Oct 30;22(1):286 [FREE Full text] [doi: 10.1186/s13054-018-2194-7] [Medline: 30373653]
18. Yu S, Leung S, Heo M, Soto G, Shah R, Gunda S, et al. Comparison of risk prediction scoring systems for ward patients: a retrospective nested case-control study. Crit Care 2014 Jun 26;18(3):R132 [FREE Full text] [doi: 10.1186/cc13947] [Medline: 24970344]
19. Subbe CP, Slater A, Menon D, Gemmell L. Validation of physiological scoring systems in the accident and emergency department. Emerg Med J 2006 Nov;23(11):841-845 [FREE Full text] [doi: 10.1136/emj.2006.035816] [Medline: 17057134]
20. Debnath S, Barnaby D, Coppa K, Makhnevich A, Kim E, Chatterjee S, Northwell COVID-19 Research Consortium. Machine learning to assist clinical decision-making during the COVID-19 pandemic. Bioelectron Med 2020;6:14 [FREE Full text] [doi: 10.1186/s42234-020-00050-8] [Medline: 32665967]
21. Ferrari D, Milic J, Tonelli R, Ghinelli F, Meschiari M, Volpi S, et al. Machine learning in predicting respiratory failure in patients with COVID-19 pneumonia-Challenges, strengths, and opportunities in a global health emergency. PLoS One 2020;15(11):e0239172 [FREE Full text] [doi: 10.1371/journal.pone.0239172] [Medline: 33180787]
22. Assaf D, Gutman Y, Neuman Y, Segal G, Amit S, Gefen-Halevi S, et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. Intern Emerg Med 2020 Nov;15(8):1435-1443 [FREE Full text] [doi: 10.1007/s11739-020-02475-0] [Medline: 32812204]
23. Haimovich A, Ravindra N, Stoytchev S, Young H, Wilson F, van Dijk D, et al. Development and Validation of the Quick COVID-19 Severity Index: A Prognostic Tool for Early Clinical Decompensation. Ann Emerg Med 2020 Oct;76(4):442-453 [FREE Full text] [doi: 10.1016/j.annemergmed.2020.07.022] [Medline: 33012378]
24. Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor Newsl 2004 Jun 01;6(1):20. [doi: 10.1145/1007730.1007735]
25. Yu H, Huang F, Lin C. Dual coordinate descent methods for logistic regression and maximum entropy models. Mach Learn 2010 Nov 11;85(1-2):41-75. [doi: 10.1007/s10994-010-5221-8]
26. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016 Apr 13 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016; San Francisco, CA p. 785-794.
27. Lemaître G, Nogueira F, Aridas C. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. J Mach Learn Res 2017;18(17):5 [FREE Full text]
28. More A. Survey of resampling techniques for improving classification performance in unbalanced datasets. arXiv Preprint posted online on August 22, 2016. [FREE Full text]
29. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, et al. Missing value estimation methods for DNA microarrays. Bioinformatics 2001 Jun;17(6):520-525. [doi: 10.1093/bioinformatics/17.6.520] [Medline: 11395428]
30. Harris SL, Harris DM. Sequential Logic Design. In: Digital Design and Computer Architecture. Burlington, MA, USA: Morgan Kaufmann; 2012.
31. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, et al. API design for machine learning software: experiences from the scikit-learn project. arXiv Preprint posted online on September 1, 2013.
32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Imputation transformer for completing missing values. Scikit-learn: Machine Learning in Python. 2011. URL: https://scikit-learn.org/stable/modules/generated/sklearn.impute.SimpleImputer.html [accessed 2020-05-08]

33. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Imputation for completing missing values using k-Nearest Neighbors. Scikit-learn: Machine Learning in Python. 2011. URL: https://scikit-learn.org/stable/modules/generated/sklearn.impute.KNNImputer.html [accessed 2020-05-08]
34. Gerry S, Birks J, Bonnici T, Watkinson P, Kirtley S, Collins G. Early warning scores for detecting deterioration in adult hospital patients: a systematic review protocol. BMJ Open 2017 Dec 03;7(12):e019268 [FREE Full text] [doi: 10.1136/bmjopen-2017-019268] [Medline: 29203508]
35. Bilben B, Grandal L, Søvik S. National Early Warning Score (NEWS) as an emergency department predictor of disease severity and 90-day survival in the acutely dyspneic patient - a prospective observational study. Scand J Trauma Resusc Emerg Med 2016 Jun 02;24:80 [FREE Full text] [doi: 10.1186/s13049-016-0273-9] [Medline: 27250249]
36. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. QJM 2001 Oct;94(10):521-526. [doi: 10.1093/qjmed/94.10.521] [Medline: 11588210]
37. Hripcsak G. Writing Arden Syntax Medical Logic Modules. Comput Biol Med 1994 Sep;24(5):331-363. [doi: 10.1016/0010-4825(94)90002-7] [Medline: 7705066]
38. Sheppard C. Tree-based Machine Learning Algorithms: Decision Trees, Random Forests, and Boosting. Scotts Valley, CA: Createspace Independent Publishing Platform; Sep 09, 2017.
39. Mani I, Zhang I. kNN approach to unbalanced data distributions: a case study involving information extraction. 2003 Presented at: International Conference on Machine Learning. Workshop on Learning from Imbalanced Datasets II; 2003; Washington, DC URL: https://www.site.uottawa.ca/~nat/Workshop2003/jzhang.pdf
40. Yu H, Huang F, Lin C. Dual coordinate descent methods for logistic regression and maximum entropy models. Mach Learn 2010 Nov 11;85(1-2):41-75. [doi: 10.1007/s10994-010-5221-8]
41. Yan L, Zhang H, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell 2020 May 14;2(5):283-288. [doi: 10.1038/s42256-020-0180-7]
42. Kumar A, Gupta PK, Srivastava A. A review of modern technologies for tackling COVID-19 pandemic. Diabetes Metab Syndr 2020;14(4):569-573 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.008] [Medline: 32413821]
43. Xu Z, Wang Z. A Risk Prediction Model for Type 2 Diabetes Based on Weighted Feature Selection of Random Forest and XGBoost Ensemble Classifier. : IEEE; 2019 Presented at: 11th International Conference on Advanced Computational Intelligence (ICACI); 2019; Guilin, China. [doi: 10.1109/icaci.2019.8778622]
44. Sharma A, Verbeke WJMI. Improving Diagnosis of Depression With XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset (n = 11,081). Front Big Data 2020 Apr 30;3:1. [doi: 10.3389/fdata.2020.00015]
45. Zabihi M, Kiranyaz S, Gabbouj M. Sepsis Prediction in Intensive Care Unit Using Ensemble of XGboost Models. : IEEE; 2019 Sep Presented at: Computing in Cardiology Conference (CinC); 2019; Singapore, Singapore URL: https://ieeexplore.ieee.org/document/9005564 [doi: 10.22489/cinc.2019.238]
46. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, China Medical Treatment Expert Group for COVID-19. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA Intern Med 2020 Aug 01;180(8):1081-1089 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2033] [Medline: 32396163]
47. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;13:8-17 [FREE Full text] [doi: 10.1016/j.csbj.2014.11.005] [Medline: 25750696]
48. Liu NT, Salinas J. Machine Learning for Predicting Outcomes in Trauma. Shock 2017 Nov;48(5):504-510. [doi: 10.1097/SHK.0000000000000898] [Medline: 28498299]
49. Ferroni P, Zanzotto F, Riondino S, Scarpato N, Guadagni F, Roselli M. Breast Cancer Prognosis Using a Machine Learning Approach. Cancers (Basel) 2019 Mar 07;11(3):328 [FREE Full text] [doi: 10.3390/cancers11030328] [Medline: 30866535]
50. Xu S, Li Y. Beware of the second wave of COVID-19. Lancet 2020 Apr 25;395(10233):1321-1322 [FREE Full text] [doi: 10.1016/S0140-6736(20)30845-X] [Medline: 32277876]
51. Bibkov B, Bibkov A. Communication on COVID-19 to community - measures to prevent a second wave of epidemic. OSF Preprints. Preprint published online on March 24, 2020 [FREE Full text] [doi: 10.31219/osf.io/ea9jm]
52. Strzelecki A. The second worldwide wave of interest in coronavirus since the COVID-19 outbreaks in South Korea, Italy and Iran: A Google Trends study. Brain Behav Immun 2020 Aug;88:950-951 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.042] [Medline: 32311493]

Abbreviations

AUC: area under the curve
AUCPR: area under the curve of the precision-recall curve
AUCROC: area under the curve of the receiver operating characteristic curve
ED: emergency department
EHR: electronic health record
ESI: Emergency Severity Index

ICU: intensive care unit
IMV: invasive mechanical ventilation
MEWS: Modified Early Warning Score
PR: precision-recall
ROC: receiver operating characteristic
SMOTE: synthetic minority oversampling
SMOTEENN: oversampling using SMOTE and cleaning using edited nearest neighbors

Edited by G Eysenbach; submitted 10.09.20; peer-reviewed by Y Xiang, S Nagavally, M Elbattah; comments to author 30.11.20; revised version received 18.12.20; accepted 18.01.21; published 10.02.21.

Please cite as:
Bolourani S, Brenner M, Wang P, McGinn T, Hirsch JS, Barnaby D, Zanos TP, Northwell COVID-19 Research Consortium
A Machine Learning Prediction Model of Respiratory Failure Within 48 Hours of Patient Admission for COVID-19: Model Development and Validation
J Med Internet Res 2021;23(2):e24246
URL: http://www.jmir.org/2021/2/e24246/
doi:10.2196/24246
PMID:33476281

Original Paper

Consuming Information Related to COVID-19 on Social Media Among Older Adults and Its Association With Anxiety, Social Trust in Information, and COVID-Safe Behaviors: Cross-sectional Telephone Survey

Frankie Ho Chun Wong[1], BSSc, MSc; Tianyin Liu[1], BSocSc, PhD; Dara Kiu Yi Leung[1], BSc, MSc, PhD; Anna Y Zhang[1], BSocSc, MSW, PhD; Walker Siu Hong Au[1], BSc, MSc; Wai Wai Kwok[1], BSc, MPsy; Angie K Y Shum[1], BA, MClinPsych; Gloria Hoi Yan Wong[1], BA, MA, PhD; Terry Yat-Sang Lum[1,2], BSocSc, MSW, PhD

[1]Department of Social Work and Social Administration, The University of Hong Kong, Pokfulam, Hong Kong
[2]Sau Po Centre on Aging, The University of Hong Kong, Pokfulam, Hong Kong

Corresponding Author:
Terry Yat-Sang Lum, BSocSc, MSW, PhD
Department of Social Work and Social Administration
The University of Hong Kong
CJT-514
The University of Hong Kong
Pokfulam
Hong Kong
Phone: 852 39178569
Email: tlum@hku.hk

Abstract

Background: COVID-19-related information on social media is overabundant and sometimes questionable, resulting in an "infodemic" during the pandemic. While previous studies suggest social media usage increases the risk of developing anxiety symptoms, how induced anxiety affects attitudes and behaviors is less discussed, let alone during a global pandemic. Little is known about the relationship between older adults using social media during a pandemic and their anxiety, their attitudes toward social trust in information, and behaviors to avoid contracting COVID-19.

Objective: The goal of this study was to investigate the associations between using social media for COVID-19-related information and anxiety symptoms as well as the mediation effect of anxiety symptoms on social trust in information and COVID-safe behaviors among older adults.

Methods: A cross-sectional telephone survey was conducted in Hong Kong between May and August 2020. A rapid warm-call protocol was developed to train social workers and volunteers from participant nongovernmental organizations to conduct the telephone surveys. Questions related to COVID-safe behaviors, social trust in information, social media use, anxiety and depressive symptoms, and sociodemographic information were asked. The number of confirmed COVID-19 cases at the community level was used to account for the risk of contracting COVID-19. Ordinary least squares regressions examined the associations between social media use and anxiety symptoms, and how they were associated with social trust in information and COVID-safe behaviors. Structural equation modeling further mapped out these relationships to identify the mediation effects of anxiety symptoms.

Results: This study collected information regarding 3421 adults aged 60 years and older. Use of social media for COVID-19-related information was associated with more anxiety symptoms and lower social trust in information but had no significant relationship with COVID-safe behaviors. Anxiety symptoms predicted lower social trust in information and higher COVID-safe behaviors. Lower social trust in information was predicted by using social media for COVID-19 information, mediated by anxiety symptoms, while no mediation effect was found for COVID-safe behaviors.

Conclusions: Older adults who rely on social media for COVID-19-related information exhibited more anxiety symptoms, while showing mixed effects on attitudes and behaviors. Social trust in information may be challenged by unverified and contradictory information online. The negligible impact on COVID-safe behaviors suggested that social media may have caused more confusion than consolidating a consistent effort against the pandemic. Media literacy education is recommended to promote critical evaluation of COVID-19-related information and responsible sharing among older adults.

(J Med Internet Res 2021;23(2):e26570) doi:10.2196/26570

KEYWORDS

COVID-19; anxiety; social media; infodemic; Hong Kong

Introduction

The COVID-19 pandemic is the first in history in which technology and social media have been used on a massive scale to keep people safe, informed, productive, and connected [1]. Although older adults are arguably less tech-savvy and do not use social media as often as younger adults, previous studies suggest that using social media for communication may help maintain relationships and prevent loneliness among older adults while having the potential to enhance health-related knowledge through information seeking [2,3]. However, the overabundance of information, especially misinformation about the COVID-19 pandemic because of its novelty and associated uncertainty, is fueling an infodemic [4]. Since the pandemic is likely to be enduring, its prolonged impact on mental health is anticipated, rendering it an important research and practice priority [5]. Therefore, effort should be made to understand the role of social media use in mental health among older adults during the COVID-19 pandemic.

Exposure to mass media in crises has been associated with anxiety symptoms and distress [6,7]. Meanwhile, the use of social media presents a higher risk of having anxiety disorder [8], where passive information consumption such as reading the news is associated with stronger anxiety symptoms [9,10]. The characteristics of social media and the way older adults consume it may further facilitate the spread of anxiety during a public health crisis such as the COVID-19 pandemic. Most recent evidence suggests an association between social media use and increased anxiety symptoms during the pandemic [11-13]. Researchers suggest that in the epicenter, Wuhan, China, this association could be a result of the fear inflicted by misinformation circulating online, while social media amplified widespread nervousness and worry [14]. Previous experiences during the Korean Middle East Respiratory Syndrome (MERS) outbreak also demonstrated the positive relationship between social media exposure and higher perceived public health risks [15]. Therefore, reliance on social media for acquiring information related to the COVID-19 pandemic introduces a higher chance of being overwhelmed by unverified and contradictory messages that promote the vicious cycle of anxiety.

During the COVID-19 pandemic, adults aged 60 years or older in Hong Kong had a significantly higher risk of exhibiting anxiety symptoms compared to the general population [16]. Challenges are presented to older adults, who may respond to the pandemic and social media differently. Contrary to the primarily unidirectional communication process in traditional media, interactive social media platforms allow users to engage in different modes of health communication [17]. A systematic review showed that previous research seldom distinguished different purposes of social media use among older adults [18]. Active usage, such as sharing or posting personal content, is different in nature from passive usage, such as browsing or seeking information that requires less effort and no need to communicate a self-concept with others in the virtual world [9]. While digital technologies designed for older adults typically stimulate passive usage [2], digital-savvy older adults are developing distinct ways of adopting internet usage, substituting online sources for traditional media for news and information consumption [19]. Passive usage is common among older adults in Hong Kong, of whom 68% use the internet for reading news and 71% listen to or watch multimedia content [20]. While their one-way consumption of information and the social media environment suggest a higher risk of having anxiety symptoms [9,10], more evidence is needed to understand whether using social media for COVID-19 information has a negative impact on older adults during the pandemic.

Worries over physical health are a major source of anxiety-related concern among older adults, where the cognitive phenomenon of worrying is central to their experience [21]. The higher the health risk, the more likely older adults feel anxious. These health-related concerns could be intensified by COVID-19-related media consumption portraying the virus as particularly harmful to the older population [22]. Inconsistent and overwhelming information available on social media exacerbates worries about the pandemic [23]. Therefore, exposure to COVID-19-related information on social media during the pandemic may induce anxiety symptoms [23-25]. It is hypothesized that older adults who use social media for COVID-19-related information will exhibit more anxiety symptoms.

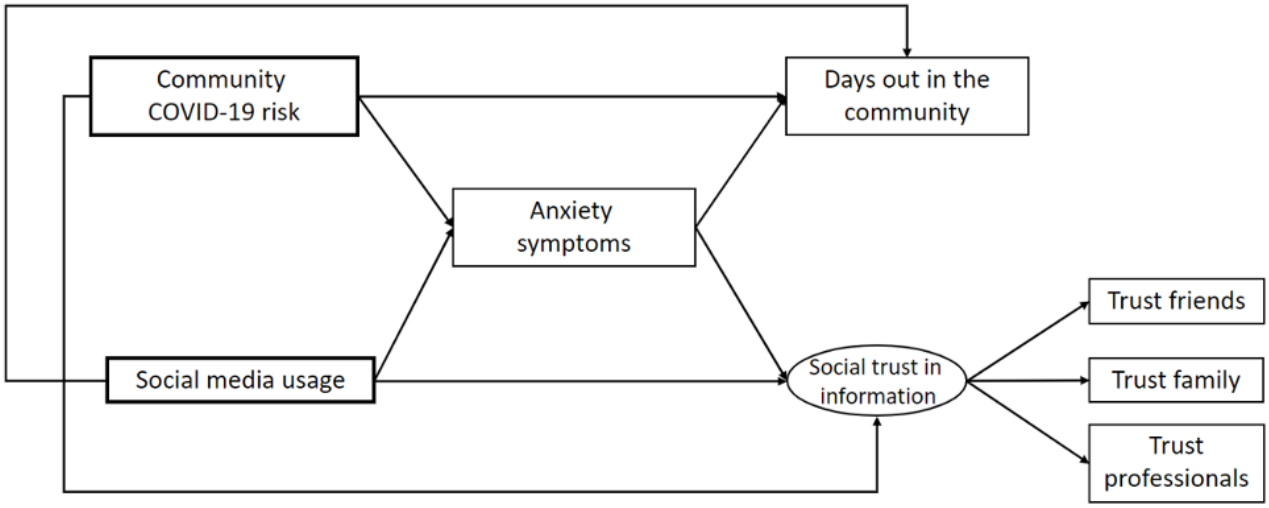
Anxiety may contribute to pandemic-related attitudes. One of the impacts of anxiety on attitudes is its association with social distrust [26]. Recent research in the context of COVID-19 demonstrates that using social media as an information source is associated with conspiracy beliefs [27]. Anxious about contradictory information online, older adults using social media may show more disbelief toward what is told. As a result, the attitudes in trusting the information about COVID-19 from the social circles around these older adults may be affected. While trust in information may also affect how much older adults use a particular media platform, this study is particularly interested in how the infodemic may affect older adults. Especially when the literature suggested that information on social media may prompt anxiety, the implication of investigating the mediation effect from using social media for COVID-19 information on social trust in information is to unveil the effects of media usage and the ways older adults are getting to know the world. It is hypothesized that older adults for whom social media is their main source of COVID-19-related information have lower social trust in information, mediated by anxiety symptoms.

While the information on social media mediates personal mitigation strategies, anxiety may prompt COVID-safe behaviors, including handwashing and social distancing [28]. Meanwhile, excessive anxiety may lead to maladaptive COVID-safe behaviors, such as panic purchasing, or

nonadherence to public health recommendations [29]. Given the immense public health risk, anxious older adults should be encouraged to adopt more COVID-safe behaviors. Wearing face masks, frequent hand hygiene, and social distancing by staying home were three COVID-safe behaviors suggested by infectious disease experts in Hong Kong [30]. Both recommendations of wearing face masks and frequent hand hygiene were widely followed by residents. For example, 98.8% of all residents wore face masks in early 2020 before it was mandated by the government [31]. However, the social distancing recommendation was observed by fewer people for various reasons (eg, work, shopping, exercise, and meals with friends). Therefore, we chose social distancing, measured by the number of days each week older adults ventured into the community, as a proxy for COVID-safe behavior in this study. It is hypothesized that older adults who use social media as their source of COVID-19-related information will reduce social contact by reducing time spent in the community, mediated by anxiety symptoms.

Anxiety symptoms, social trust in information, and time spent in the community may be influenced by the objective risk of contracting COVID-19. In order to distinguish the effects of the actual risk from the effects of the perceptions acquired via social media, the hypothesized theoretical framework put the risk of contracting COVID-19 as another independent factor (see Figure 1).

Figure 1. Theoretical framework.



The specific focus of this study was to investigate the effects of consuming public health–related information via social media on anxiety among the older population during a public health crisis. Some of the most recent studies investigating the impact of information sources on confidence in coping with COVID-19 [32], behaviors [33], and beliefs about health-related information [34] have primarily used internet-based questionnaires to survey their participants. Since, in this study, we were interested in evaluating older people, some of whom may be less engaged in digital communication, an online survey would present difficulties in reaching them and could create sampling bias. Therefore, a telephone survey was chosen to capture a more diverse sample. To address the gap in the current literature, this study investigated how using social media for COVID-19 information among older adults during the pandemic was associated with anxiety symptoms and how the two contribute to social trust in information and COVID-safe behaviors.

Methods

Study Design and Sampling

A cross-sectional telephone survey was conducted during the COVID-19 outbreak in Hong Kong between May and August 2020; participants included adults aged 60 years and older using services provided by community elderly centers and community mental wellness centers. The former centers provide active aging programs and social support for community-dwelling older adults. The latter centers run programs to promote mental wellness in the community and provide psychosocial support for community-dwelling residents of all ages with mental health challenges. The research team developed a short survey instrument and trained center-based social workers and volunteers to conduct the survey. We called a total of 3550 members and completed 3421 interviews, giving a success rate of 96.4%. All respondents had no prior history of COVID-19.

Data Collection

Data were collected through telephone interviews. Each interview lasted for about 10 to 15 minutes. The questionnaire was tested and revised before full-scale implementation. Qualified clinical psychologists and the researchers of this study designed the drafted questionnaire, which was reviewed and pretested by frontline social workers. Questions that were identified as challenging for older people with limited education to understand were revised and simplified. The process was repeated for two rounds before large-scale implementation. Interviewers were trained to make phone calls to respondents using a standardized, rapid, warm-call protocol. The protocol started with a warm call; that is, getting in touch with service recipients with known personal information. After greeting them and asking for their needs amid the pandemic, the questionnaire was fielded. Frontline social workers followed up with participants who demonstrated mental health challenges. This

study was approved by the Human Research Ethics Committee of the University of Hong Kong (reference No. EA2003001[A]).

Measurements

The major dependent variables of the study were anxiety, social trust in information, and COVID-safe behaviors. Anxiety was measured using the validated Chinese version of the 2-item Generalized Anxiety Disorder (GAD-2) scale, which ranged from 0 to 6, where a GAD-2 score of 3 or more indicated a high possibility of anxiety [35,36]. For social trust in information, respondents were asked to rate the extent to which they believed the COVID-19-related information that was shared from people in their social circles, including family, friends, and professionals (eg, social workers and general practitioners), on a scale from 1 (total distrust) to 5 (complete trust). The scale included a *not applicable* option that resulted in only 1480 respondents completing all three questions. Social trust in information was computed as the average of the three items for ordinary least squares (OLS) regressions (Cronbach α=.69) and as a latent variable of the three items in structural equation modeling (SEM). Only responses with complete answers in all three items were compiled. By treating social trust in information as missing when the respondents selected *not applicable* in any of the three components, the valid number of responses of social trust in information in OLS regressions was reduced to 1480. In the SEM, missing data were treated with full information maximum likelihood (FIML) estimation. The sample size for the SEM was 3415. COVID-safe behaviors were proxied by social distancing, measured by the older adults' responses to how many days each week they ventured into the community.

The major independent variable was social media usage for COVID-19-related information, which was measured with three categories: (1) no usage, (2) social media used for COVID-19-related information, and (3) social media used as the main source of COVID-19-related information. Usage of social media for COVID-19-related information was recoded from a survey item on the trust level of media sources. The trust levels of social media and traditional media were captured in a 5-point scale, ranging from 1 (total distrust) to 5 (complete trust). The scale included a *not applicable* option. Only respondents using social media for COVID-19-related information were prompted to answer the question about trust level of social media, while nonusers picked the *not applicable* option. To measure any use of social media for COVID-19-related information, valid responses on social media trust were, therefore, recoded as *used social media for COVID-19-related information*. The nonusers who picked *not applicable* were considered to have no usage. The main information source was captured by a dichotomous variable of *traditional media* or *social media*, where the *social media* responses were recoded into *social media used as the main source of COVID-19-related information*.

The control variables include the risk of contracting COVID-19, depressive symptoms, and demographics. The risk of contracting COVID-19 was measured by the number of all confirmed cases in the week of data collection in the district in which the respondent lived. This information was obtained from the government daily report of newly confirmed cases. This measurement is geographically sensitive and reflected the risk of anyone contracting COVID-19, older adults included, who ventured into their neighborhood. Depression was measured using the validated Chinese version of the 2-item Patient Health Questionnaire (PHQ-2), with scores ranging from 0 and 6, where a PHQ-2 score of 3 or more indicates a high possibility of depression [37,38]. Demographics collected included respondents' age in years, gender (ie, male or female), district of residency, and membership of service unit (ie, elderly center or mental wellness center). Respondents from mental wellness centers, who had a history of mental challenges, were included in the sample. A dummy variable was added to statistical models to control for its effects on mental health status and to investigate any difference in using social media for COVID-19 information.

Statistical Analysis

Descriptive statistics of all variables were computed and reported as appropriate. Social media usage categories were entered into the statistical models as dummy variables of *used social media for COVID-19 information* and *used social media as the main COVID-19 information source*, with *no usage* as the reference group. Multivariate OLS regressions were applied to investigate the effect of using social media for COVID-19 information on anxiety symptoms as well as the effects of using social media for COVID-19 information and anxiety symptoms on time spent in the community and social trust in information. The theoretical model was then examined by SEM to test the mediation effect of anxiety symptoms. Social trust in information was treated as a latent construct with the three trust items. The first level of the SEM predicted anxiety symptoms with depressive symptoms, using social media for COVID-19 information, and demographics. The second level predicted time in the community and social trust in information with using social media for COVID-19 information, anxiety symptoms, and demographics. The FIML approach, which estimates parameters and standard errors directly with all data but does not impute missing data, was employed to account for the missing data in SEM. FIML generates relatively unbiased estimates and the least convergence failures compared to other methods, such as listwise or pairwise deletion and multiple imputations in SEM [39]. It is also less affected by nonnormal missing data and data distribution shape [40].

All data were consolidated and analyzed with SPSS software, version 26 (IBM Corp), after the removal of personal identifying information. SEM was conducted using the R package *lavaan* (The R Foundation) [41].

Results

Table 1 shows the respondents' characteristics. Their average age was 76 years (SD 8.9), 25.4% (869/3418) were male, and 77.9% (2666/3421) were recruited from elderly centers. The average number of COVID-19 cases in a community was 25.7 (SD 27.5), ranging from 0 to 135. Depression and anxiety were not prevalent: the average PHQ-2 score was 0.83 (SD 1.3) and the average GAD-2 score was 0.74 (SD 1.2). While 8.7% (298/3421) of the respondents exhibited a risk of having depression (ie, PHQ-2 score ≥3), 7.0% (241/3421) of the

respondents exhibited a risk of having anxiety disorder (ie, GAD-2 score ≥3). Out of 3421 respondents, 1399 (40.9%) used social media to obtain COVID-19-related information and 203 (5.9%) used social media as their main source of COVID-19 information. The respondents had spent some time outside the home in the community on an average of 4 days (SD 2.4) in a week. In terms of their attitudes, the respondents had moderate to high levels of social trust in information (family: mean 4.4, SD 0.83; friends: mean 3.6, SD 0.96; professionals: mean 4.5, SD 0.74).

Compared to respondents recruited from mental wellness centers, respondents from elderly centers showed a different profile in their demographics and mental health status, but not in their usage of social media. Respondents from elderly centers were, on average, 4.4 years older (t_{3419}=12.0, *P*<.001), and more of them were male (elderly centers: 703/2663, 26.4%; mental wellness centers: 166/755, 22.0%; χ^2_1=6.0, *P*=.01). They also showed fewer depressive symptoms as measured by the PHQ-2 (elderly centers: mean 0.69, SD 1.18; mental wellness centers: mean 1.33, SD 1.53; t_{3416}=12.3, *P*<.001) and fewer anxiety symptoms as measured by the GAD-2 (elderly centers: mean 0.61, SD 1.13; mental wellness centers: mean 1.22, SD 1.43; t_{3386}=12.1, *P*<.001). Nevertheless, there was no significant difference in their usage of social media (t_{3419}=1.35, *P*=.18), which was the main independent variable. To account for the potential bias from a different mental health profile, the nature of the service was treated as a control variable in the subsequent analysis.

Table 1. Respondent characteristics.

Variable	Value (N=3421)
Demographics	
Age (years), mean (SD)	76.0 (8.9)
Gender (male) (n=3418), n (%)	869 (25.4)
Service nature (aged care), n (%)	2666 (77.9)
Community COVID-19 risk, mean (SD)	
Weekly number of COVID-19 cases in district	25.7 (27.5)
Psychological well-being (range 0-6), mean (SD)	
2-item Patient Health Questionnaire (n=3418)	0.83 (1.3)
2-item Generalized Anxiety Disorder scale (n=3388)	0.74 (1.2)
Trust in media (range 1-5), mean (SD)	
Traditional media (n=3335)	4.27 (0.88)
Social media (n=1399)	3.18 (1.1)
Using social media for COVID-19 information	
Used social media for COVID-19 information, n (%)	1399 (40.9)
Social media as the main source of COVID-19 information, n (%)	203 (5.9)
Days out in the community per week (n=3161), mean (SD)	4.1 (2.4)
Social trust in COVID-19 information (range 1-5), mean (SD)	
Family (n=2620)	4.4 (0.83)
Friends (n=2023)	3.6 (0.96)
Professionals (n=2150)	4.5 (0.74)

Table 2 shows the results of the OLS regressions. After controlling for demographics and depressive symptoms, the category *used social media for COVID-19-related information* was not associated with anxiety symptoms but *used social media as the main source of COVID-19 information* did (B=0.18, *P*=.003). Using social media as the main source for COVID-19-related information was associated with more anxiety symptoms. Although social media use did not predict time spent in the community, it was associated with lower social trust in information. Both levels of usage—*used social media for COVID-19-related information* (B=–0.11, *P*=.005) and *used social media as the main source of COVID-19 information* (B=–0.30, *P*<.001)—showed a negative association. Anxiety symptoms were also negatively associated with time spent in the community and social trust in information (days out: B=–0.08, *P*=.02; social trust: B=–0.04, *P*=.02). Community COVID-19 risk was not significantly associated with anxiety symptoms (B=0.001, *P*=.69) but predicted less time spent in the community (B=–0.01, *P*<.001) and higher social trust in information (B=0.003, *P*<.001). Aged care respondents showed fewer anxiety symptoms (B=–0.17, *P*<.001) and higher social trust in information (B=0.23, *P*<.001), while older respondents spent the least amount of time in the community (B=0.03, *P*<.001). Male respondents were less anxious (B=–0.09, *P*=.004), spent more days out in the community (B=0.24,

P=.02), and had lower social trust in information (B=–0.10, *P*=.02).

Table 2. Results from ordinary least squares regressions predicting anxiety symptoms, days out in the community, and social trust in information.

Variable	GAD-2[a] scale (n=3384)		Days out in the community (n=3126)		Social trust in information (n=1480)	
	B[b] (95% CI)	*P* value	B (95% CI)	*P* value	B (95% CI)	*P* value
Gender (male)	–0.09 (–0.15 to –0.03)	.004	0.24 (0.04 to 0.43)	.02	–0.10 (–0.18 to –0.02)	.02
Age (years)	–0.002 (0)	.19	–0.03 (–0.04 to –0.02)	<.001	–0.001 (–0.01 to 0)	.80
Community COVID-19 risk	0.001 (0)	.07	–0.01 (–0.01 to 0)	<.001	0.003 (0)	<.001
Service nature (aged care)	–0.17 (–0.24 to –0.10)	<.001	0.04 (–0.17 to 0.26)	.68	0.23 (0.14 to 0.32)	<.001
Depressive symptoms	0.71 (0.69 to 0.73)	<.001	N/A[c]	N/A	N/A	N/A
Used social media for COVID-19 information	0.002 (0 to 0.10)	.95	0.02 (–0.17 to 0.22)	.81	–0.11 (–0.19 to –0.03)	.005
Used social media as the main COVID-19 information source	0.18 (0.06 to 0.30)	.003	0 (–0.36 to 0.36)	.99	–0.30 (–0.43 to –0.16)	<.001
Anxiety symptoms	N/A	N/A	–0.08 (–0.15 to –0.01)	.02	–0.04 (–0.07 to –0.01)	.02

[a]GAD-2: 2-item Generalized Anxiety Disorder.

[b]Unstandardized coefficient.

[c]N/A: not applicable. The 2-item Patient Health Questionnaire (PHQ-2) was used for controlling the comorbidity with anxiety in the equation regressing GAD-2; there are no PHQ-2 values because the PHQ-2 was not included in the equations.

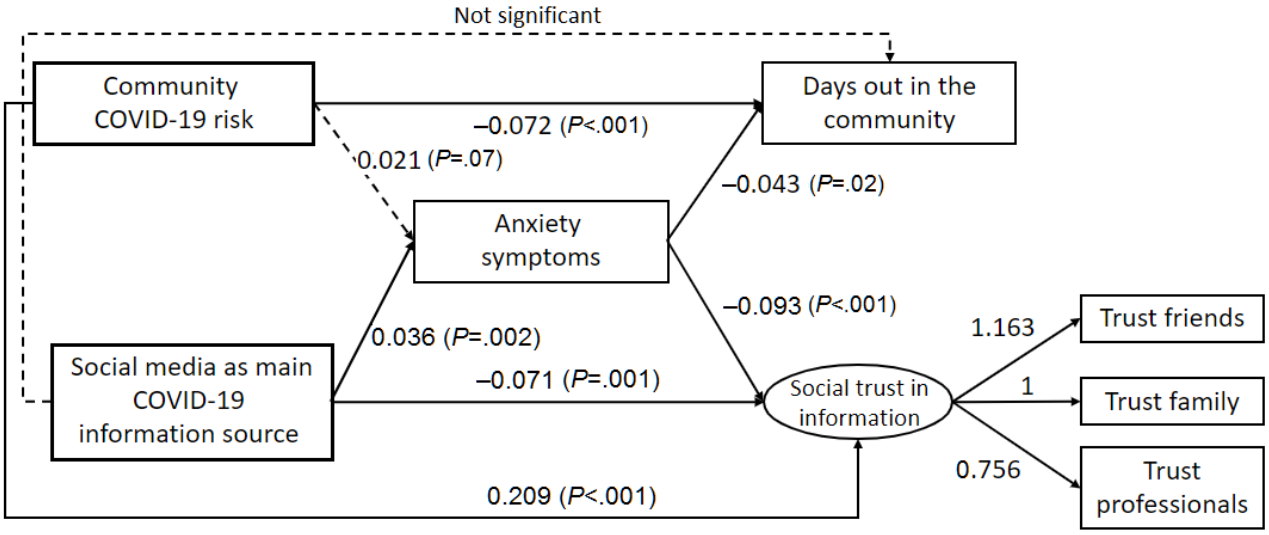
The results of SEM analysis (see Table 3 and Figure 2), modeling the mediation effect of anxiety symptoms, showed a good fit (χ^2_{19}=48.0, *P*<.001; Comparative Fit Index [CFI]=0.993; Tucker-Lewis Index [TLI]=0.985; root mean square error of approximation [RMSEA]=0.021; standardized root mean residual=0.014) and coincided with most findings in the OLS regressions. Since a lower level of social media use for COVID-19-related information was not associated with anxiety symptoms in the OLS regressions, the category was collapsed with *no social media usage* in the SEM. Using social media as the main source of COVID-19-related information predicted more anxiety symptoms (SEM coefficient=0.036, *P*=.002). Anxiety, meanwhile, was associated with lower social trust in information (SEM coefficient=–0.093, *P*<.001) and less time spent in the community (SEM coefficient=–0.043, *P*=.02). Using social media as the main source of COVID-19 information had a direct association with lower social trust in information (SEM coefficient=–0.071, *P*=.001), which means relying on social media for COVID-19 information was a significant predictor of social trust in information. Moreover, an indirect effect of using social media for COVID-19 information was mediated by anxiety (total effect=–0.074, *P*=.001). However, using social media as the main source of COVID-19 information was not associated with time spent in the community (SEM coefficient=–0.001, *P*=.97). No mediation effect was detected either (total effect=–0.002, *P*=.90). Sensitivity checking using social trust in information and time spent in the community to predict anxiety symptoms in SEM showed no significant associations (CFI=0.988; TLI=0.976; RMSEA=0.027), suggesting that anxiety was a valid independent factor in the second level of the original SEM model.

Table 3. Standardized parameter estimates and their standard errors of structural equation modeling (N=3415) treated with full information maximum likelihood estimation.

Equation model or regression	Standardized parameter estimate	SE	*P* value
Measurement model			
Social trust in information measured by trust in family	1	N/A[a]	N/A
Social trust in information measured by trust in friends	1.163	0.064	<.001
Social trust in information measured by trust in professionals	0.756	0.041	<.001
Regression			
Anxiety symptoms measured by the following:			
Age (years)	–0.017	0.002	.15
Gender (male)	–0.032	0.032	.004
Service nature (aged care)	–0.057	0.036	<.001
Community COVID-19 risk	0.021	0.001	.07
Depressive symptoms	0.740	0.011	<.001
Social media as the main COVID-19 information source	0.036	0.060	.002
Social trust in information measured by the following:			
Age (years)	0.106	0.001	<.001
Gender (male)	–0.062	0.030	.007
Service nature (aged care)	0.138	0.032	<.001
Community COVID-19 risk	0.209	0	<.001
Social media as the main COVID-19 information source	–0.071	0.052	.001
Anxiety symptoms	–0.093	0.011	<.001
Days out in the community measured by the following:			
Age (years)	–0.102	0.005	<.001
Gender (male)	0.042	0.097	.02
Service nature (aged care)	0.008	0.108	.69
Community COVID-19 risk	–0.072	0.002	<.001
Social media as the main COVID-19 information source	–0.001	0.176	.97
Anxiety symptoms	–0.043	0.035	.02

[a]N/A: not applicable; this is the reference variable.

Figure 2. Structural equation modeling. Standardized parameter estimates are reported.



Discussion

Principal Findings

Our findings are not only consistent with previous studies on the association between social media use and anxiety [8,9], especially during the COVID-19 pandemic [11-13], but they also identified the major type of social media use among older adults (ie, consuming information) as predicting more anxiety symptoms. Both the OLS regression and SEM results demonstrated the association between using social media for COVID-19-related information and more anxiety symptoms among older adults. On the one hand, the ecology of social media, characterized by information overload and sensationalizing of information [42,43], could have amplified older users' nervousness and perceptions of the risks of contracting COVID-19. The inconsistency in unverified information on social media exacerbates this anxiety. On the other hand, older people's passive usage of social media suggests an intensified anxiety [9,10]. Since mental well-being decreases when social media users fail to acquire direct social interaction to satisfy their relatedness needs [44,45], passive usage may intensify unmet needs resulting from social isolation during the pandemic, thus prompting anxiety symptoms.

The SEM results suggested that anxiety symptoms mediated the effect of using social media for COVID-19 information on social trust in information. It demonstrated the two pathways of the associations between using social media and attitudes in trusting information. The proliferation of unverified and emotional information may drive anxiety and lower social trust in information among social media users. Algorithms in social media platforms prioritize sensational information [43]. User attention is drawn to the emotions induced while diminishing the ability to critically evaluate the accuracy of information, promoting the impulsive spread of unverified misinformation [46]. As a result, social media users may face contradictory information on social media, which may also be inconsistent with the information acquired from family, friends, or even health professionals. Not only would they become anxious about what to believe, but social trust in information was also undermined by reading conflicting claims. Another problem brought by social media use is the overwhelming amount of information overloading users as a result of uncontrolled news consumption [42]. On popular social media platforms, such as Facebook and Twitter, an infinite scroll feeds content to users without a "stopping cue" [47], while algorithms tailor content that suits user preference [43]. The environment in social media suggests that users are encouraged to stay on their sites as long as possible, consuming more information than intended. Overload is further magnified by the challenge to process the unverified, anonymous, and overwhelmingly subjective news on social media [48]. Consequently, it is difficult to digest all the information flooding into the social media feed because of both its enormous quantity and questionable quality.

Differences were observed from various demographic profiles as well. Older age was associated with spending fewer days venturing into the community, which was likely related to the lowered mobility in aged respondents. Male respondents were less anxious and more critical of COVID-19 information from people around them, and they went into the community more often. The results may have demonstrated a more authoritative gender role in older East Asian men. Respondents from mental wellness centers showed more anxiety symptoms and lower social trust in information. This suggested that a history of mental health challenges may exacerbate the effects of the infodemic.

Our results demonstrated that the infodemic may have caused more confusion than promoting COVID-safe behaviors. While older adults' usage of social media for COVID-19 information showed no association, the number of COVID-19 cases in the community and anxiety symptoms predicted less time spent in the community, suggesting that older adults adopted COVID-safe behaviors when the risk was high and when they became anxious. COVID-safe behaviors may have been further encouraged by the surge of social distancing and stay-home advice when the number of confirmed cases grew. However, these messages might not be transmitted equally to all social media users. Health-related misinformation proliferated well before the pandemic, facilitated by "information silos" and

"echo chambers" in social media [49]. The clustering of users in social media means that both useful information and harmful misinformation may not reach a mass audience but is amplified within social bubbles. Different stories could be proliferating among different user subgroups, resulting in the uneven effects of consuming information on social media regarding COVID-safe behaviors.

A greater understanding of how anxiety affects behaviors could inform ways to communicate COVID-19-related information to older adults to promote appropriate COVID-safe behaviors and social trust in information and contribute to the wider public health effort. On the one hand, our results suggested that anxiety encouraged older adults to avoid social contact. Nevertheless, more research is warranted to understand if this is rational risk aversion or panic-induced maladaptive behavior. On the other hand, while it is normal to be anxious in the face of a global pandemic, future research should also focus on the management and mitigation of anxiety during a prolonged pandemic so that older adults' trust in the people around them, who may provide the most adequate and timely support, is less likely to be undermined.

Although this study suggests some adverse effects of using social media, it does not mean that older adults should avoid using it. On the contrary, it is necessary to address these problems and identify the potential benefit of using social media to promote adequate COVID-safe behavior and reduce loneliness amid social distancing. Digital skills alone may no longer be sufficient to enable older adults to navigate the complex contemporary media environment. Media literacy among older adults should be advocated to enable them to critically analyze and interpret information. Guidance should be provided to promote healthy coping behaviors, such as verifying suspicious information and searching for credible information sources in response to the infodemic in social media. Advice on responsible social media use is essential for building a constructive social media environment for older adults. Media literate older adults should understand the consequences of spreading misinformation and refrain from sharing unverified information. By empowering older adults to distinguish and reject misinformation, the infodemic may be alleviated if the transmission of misinformation is curbed in its early stages. Anxiety may be reduced when older adults acquire agency while mutually building a healthy social media environment.

Limitations

Since the rapid warm-call protocol was designed to conduct interviews quickly with an extensive reach, our questions were designed to be simple and concise. While the data collection period spanned across 4 months, the study was cross-sectional in nature, which could not allow it to demonstrate causal relationships between use of social media for COVID-19, social trust, and anxiety. The self-reported survey design may also suffer from data inaccuracy as a result of issues including social desirability bias or loss of memory. Although a measurement for social media usage for COVID-19 information was constructed for this study, the sensitivity of the measurement may have been compromised for the conciseness of the survey. The range in measuring *used social media for COVID-19 information* may be too wide to fully capture its pro rata effects. On top of that, the frequency of media use is yet to be addressed. The frequency or time spent on social media may provide greater information about the influence of media usage patterns on anxiety, attitudes, and behaviors. Our scale did not capture the variance in the amount of time exposed to social media, only its ratio. Therefore, the study may underestimate the effect of using social media for COVID-19 information. Meanwhile, only one COVID-safe behavior—staying home to avoid social contact—was measured. Potential bias may occur when other factors contribute to older adults staying at home, such as the closure of businesses and social services, that were unrelated to using social media for COVID-19 information and anxiety.

Our sample, recruited from both elderly and mental wellness centers, was not a representative sample of the city. The findings may not be generalized to the whole population. Without support from social service units, the unsampled cluster of older adults could have experienced a greater challenge in navigating the pandemic. They might exhibit greater anxiety from not receiving adequate social support, and social media might be a more significant source of information about the COVID-19 pandemic.

Conclusions

This study suggests that using social media for COVID-19 information by older adults in Hong Kong during the COVID-19 pandemic was associated with anxiety symptoms. Meanwhile, anxiety symptoms mediated the effect from using social media for COVID-19 information on social trust in information but not COVID-safe behaviors. This suggests that the infodemic may have caused confusion when older adults digested health information, prompting anxiety among them about what and who to believe and, thus, reducing social trust in information. The fragmentized social media landscape further promoted an uneven impact of the infodemic between different groups of users. It could explain why objective risk and anxiety predicted older adults' avoidance of leaving home but using social media for COVID-19 information did not. To encourage the healthy use of social media, media literacy education should be promoted among older adults.

Acknowledgments

The authors would like to thank the social workers and volunteers for their help in recruiting study participants and collecting data, the research assistants for checking and managing data, and all participants for their contribution to this research. The Jockey Club Holistic Support Project for Elderly Mental Wellness (JC JoyAge) is funded by The Hong Kong Jockey Club Charities Trust.

Conflicts of Interest

None declared.

References

1. Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. Joint statement by WHO, UN, UNICEF, UNDP, UNESCO, UNAIDS, ITU, UN Global Pulse, and IFRC. World Health Organization. 2020 Sep 23. URL: tinyurl.com/4fg4r76x [accessed 2020-10-21]
2. Leist AK. Social media use of older adults: A mini-review. Gerontology 2013 Apr 16;59(4):378-384 [FREE Full text] [doi: 10.1159/000346818] [Medline: 23594915]
3. Vošner HB, Bobek S, Kokol P, Krečič MJ. Attitudes of active older internet users towards online social networking. Comput Human Behav 2016 Feb;55:230-241. [doi: 10.1016/j.chb.2015.09.014]
4. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
5. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: A call for action for mental health science. Lancet Psychiatry 2020 Jun 15;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
6. Niitsu K, Watanabe-Galloway S, Sayles H, Houfek J, Rice M. A pilot study of the psychological impact of the great East Japan earthquake and tsunami. J Am Psychiatr Nurses Assoc 2014 May 05;20(3):194-202. [doi: 10.1177/1078390314536615] [Medline: 24904039]
7. Collimore KC, McCabe RE, Carleton RN, Asmundson GJ. Media exposure and dimensions of anxiety sensitivity: Differential associations with PTSD symptom clusters. J Anxiety Disord 2008 Aug;22(6):1021-1028. [doi: 10.1016/j.janxdis.2007.11.002] [Medline: 18093798]
8. Vannucci A, Flannery KM, Ohannessian CM. Social media use and anxiety in emerging adults. J Affect Disord 2017;207:163-166. [doi: 10.1016/j.jad.2016.08.040] [Medline: 27723539]
9. Thorisdottir IE, Sigurvinsdottir R, Asgeirsdottir BB, Allegrante JP, Sigfusdottir ID. Active and passive social media use and symptoms of anxiety and depressed mood among Icelandic adolescents. Cyberpsychol Behav Soc Netw 2019 Aug;22(8):535-542. [doi: 10.1089/cyber.2019.0079] [Medline: 31361508]
10. Kim S, Favotto L, Halladay J, Wang L, Boyle MH, Georgiades K. Differential associations between passive and active forms of screen time and adolescent mood and anxiety disorders. Soc Psychiatry Psychiatr Epidemiol 2020 Nov 13;55(11):1469-1478. [doi: 10.1007/s00127-020-01833-9] [Medline: 32055896]
11. Chao M, Xue D, Liu T, Yang H, Hall BJ. Media use and acute psychological outcomes during COVID-19 outbreak in China. J Anxiety Disord 2020 Aug;74:102248 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102248] [Medline: 32505918]
12. Ni MY, Yang L, Leung CMC, Li N, Yao XI, Wang Y, et al. Mental health, risk factors, and social media use during the COVID-19 epidemic and cordon sanitaire among the community and health professionals in Wuhan, China: Cross-sectional survey. JMIR Ment Health 2020 May 12;7(5):e19009 [FREE Full text] [doi: 10.2196/19009] [Medline: 32365044]
13. Nekliudov NA, Blyuss O, Cheung KY, Petrou L, Genuneit J, Sushentsev N, et al. Excessive media consumption about COVID-19 is associated with increased state anxiety: Outcomes of a large online survey in Russia. J Med Internet Res 2020 Sep 11;22(9):e20955 [FREE Full text] [doi: 10.2196/20955] [Medline: 32788143]
14. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020 Apr 16;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
15. Choi D, Yoo W, Noh G, Park K. The impact of social media on risk perceptions during the MERS outbreak in South Korea. Comput Human Behav 2017 Jul;72:422-431 [FREE Full text] [doi: 10.1016/j.chb.2017.03.004] [Medline: 32288176]
16. Zhao SZ, Wong JYH, Luk TT, Wai AKC, Lam TH, Wang MP. Mental health crisis under COVID-19 pandemic in Hong Kong, China. Int J Infect Dis 2020 Nov;100:431-433 [FREE Full text] [doi: 10.1016/j.ijid.2020.09.030] [Medline: 32947051]
17. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: Implications for health communication. J Med Internet Res 2009 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]
18. Hunsaker A, Hargittai E. A review of internet use among older adults. New Media Soc 2018 Jul 16;20(10):3937-3954. [doi: 10.1177/1461444818787348]
19. Quan-Haase A, Martin K, Schreurs K. Interviews with digital seniors: ICT use in the context of everyday life. Inf Commun Soc 2016 Feb 17;19(5):691-707. [doi: 10.1080/1369118x.2016.1140217]
20. Personal computer and internet penetration. Thematic Household Survey Report No 69. Hong Kong, China: Census and Statistics Department, Hong Kong Special Administrative Region; 2020 Mar. URL: https://www.ogcio.gov.hk/en/about_us/facts/doc/householdreport2020_69.pdf [accessed 2020-10-21]
21. Diefenbach GJ, Stanley MA, Beck JG. Worry content reported by older adults with and without generalized anxiety disorder. Aging Ment Health 2001 Aug;5(3):269-274. [doi: 10.1080/13607860120065069] [Medline: 11575066]
22. Daoust JF. Elderly people and responses to COVID-19 in 27 countries. PLoS One 2020 Jul 2;15(7):e0235590 [FREE Full text] [doi: 10.1371/journal.pone.0235590] [Medline: 32614889]

23. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for managing the COVID-19 infodemic: Methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
24. Wong LP, Hung C, Alias H, Lee TS. Anxiety symptoms and preventive measures during the COVID-19 outbreak in Taiwan. BMC Psychiatry 2020 Jul 16;20(1):376 [FREE Full text] [doi: 10.1186/s12888-020-02786-8] [Medline: 32677926]
25. Ahmad AR, Murad HR. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: Online questionnaire study. J Med Internet Res 2020 May 19;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
26. Delhey J, Newton K. Who trusts?: The origins of social trust in seven societies. Eur Soc 2003 Jan;5(2):93-137. [doi: 10.1080/1461669032000072256]
27. Allington D, Duffy B, Wessely S, Dhavan N, Rubin J. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. Psychol Med 2020 Jun 09:1-7. [doi: 10.1017/s003329172000224x]
28. Lauri Korajlija A, Jokic-Begic N. COVID-19: Concerns and behaviours in Croatia. Br J Health Psychol 2020 May 16;25(4):849-855 [FREE Full text] [doi: 10.1111/bjhp.12425] [Medline: 32779816]
29. Asmundson GJ, Taylor S. How health anxiety influences responses to viral outbreaks like COVID-19: What all decision-makers, health authorities, and health care professionals need to know. J Anxiety Disord 2020 Apr;71:102211 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102211] [Medline: 32179380]
30. Health tips: Together, we fight the virus!. The Government of the Hong Kong Special Administrative Region. 2020. URL: https://www.coronavirus.gov.hk/eng/index.html#Health_Tips [accessed 2020-10-21]
31. Wong SH, Teoh JYC, Leung C, Wu WKK, Yip BHK, Wong MCS, et al. COVID-19 and public interest in face mask use. Am J Respir Crit Care Med 2020 Aug 01;202(3):453-455. [doi: 10.1164/rccm.202004-1188le]
32. Wang P, Lu W, Ko N, Chen Y, Li D, Chang Y, et al. COVID-19-related information sources and the relationship with confidence in people coping with COVID-19: Facebook survey study in Taiwan. J Med Internet Res 2020 Jun 05;22(6):e20021 [FREE Full text] [doi: 10.2196/20021] [Medline: 32490839]
33. Ali KF, Whitebridge S, Jamal MH, Alsafy M, Atkin SL. Perceptions, knowledge, and behaviors related to COVID-19 among social media users: Cross-sectional study. J Med Internet Res 2020 Sep 08;22(9):e19913 [FREE Full text] [doi: 10.2196/19913] [Medline: 32841153]
34. Ali SH, Foreman J, Tozan Y, Capasso A, Jones AM, DiClemente RJ. Trends and predictors of COVID-19 information sources and their relationship with knowledge and beliefs related to the pandemic: Nationwide cross-sectional study. JMIR Public Health Surveill 2020 Oct 08;6(4):e21071 [FREE Full text] [doi: 10.2196/21071] [Medline: 32936775]
35. Luo Z, Li Y, Hou Y, Zhang H, Liu X, Qian X, et al. Adaptation of the two-item generalized anxiety disorder scale (GAD-2) to Chinese rural population: A validation study and meta-analysis. Gen Hosp Psychiatry 2019 Sep;60:50-56. [doi: 10.1016/j.genhosppsych.2019.07.008] [Medline: 31326672]
36. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: A systematic review and diagnostic meta-analysis. Gen Hosp Psychiatry 2016 Mar;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
37. Yu X, Stewart SM, Wong PT, Lam TH. Screening for depression with the Patient Health Questionnaire-2 (PHQ-2) among the general population in Hong Kong. J Affect Disord 2011 Nov;134(1-3):444-447. [doi: 10.1016/j.jad.2011.05.007] [Medline: 21665288]
38. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res 2005 Feb;58(2):163-171. [doi: 10.1016/j.jpsychores.2004.09.006] [Medline: 15820844]
39. Enders C, Bandalos D. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Struct Equ Modeling 2001 Jul 1;8(3):430-457 [FREE Full text]
40. Enders CK. The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. Psychol Methods 2001;6(4):352-370. [doi: 10.1037/1082-989x.6.4.352]
41. Rosseel Y. lavaan: An R package for structural equation modeling. J Stat Softw 2012;48(2):1-36 [FREE Full text] [doi: 10.18637/jss.v048.i02]
42. Holton AE, Chyi HI. News and the overloaded consumer: Factors influencing information overload among news consumers. Cyberpsychol Behav Soc Netw 2012 Nov;15(11):619-624. [doi: 10.1089/cyber.2011.0610] [Medline: 23020743]
43. Bakir V, McStay A. Fake news and the economy of emotions. Digit Journal 2017 Jul 20;6(2):154-175. [doi: 10.1080/21670811.2017.1345645]
44. Hall JA, Kearney MW, Xing C. Two tests of social displacement through social media use. Inf Commun Soc 2018 Feb 02;22(10):1396-1413. [doi: 10.1080/1369118x.2018.1430162]
45. Hall JA. When is social media use social interaction? Defining mediated social interaction. New Media Soc 2016 Jul 28;20(1):162-179. [doi: 10.1177/1461444816660782]
46. Pennycook G, McPhetres J, Zhang Y, Lu J, Rand D. Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. Psychol Sci 2020 Jul;31(7):770-780 [FREE Full text] [doi: 10.1177/0956797620939054] [Medline: 32603243]

47. Noë B, Turner LD, Linden DE, Allen SM, Winkens B, Whitaker RM. Identifying indicators of smartphone addiction through user-app interaction. Comput Human Behav 2019 Oct;99:56-65 [FREE Full text] [doi: 10.1016/j.chb.2019.04.023] [Medline: 31582873]
48. Pentina I, Tarafdar M. From "information" to "knowing": Exploring the role of social media in contemporary news consumption. Comput Human Behav 2014 Jun;35:211-223. [doi: 10.1016/j.chb.2014.02.045]
49. Chou WS, Oh A, Klein WMP. Addressing health-related misinformation on social media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]

Abbreviations

CFI: Comparative Fit Index
FIML: full information maximum likelihood
GAD-2: 2-item Generalized Anxiety Disorder
MERS: Middle East Respiratory Syndrome
OLS: ordinary least squares
PHQ-2: 2-item Patient Health Questionnaire
RMSEA: root mean square error of approximation
SEM: structural equation modeling
TLI: Tucker-Lewis Index

Edited by R Kukafka, C Basch; submitted 17.12.20; peer-reviewed by Z Ma, MA Bahrami, A Jolliff; comments to author 11.01.21; revised version received 29.01.21; accepted 31.01.21; published 11.02.21.

Please cite as:
Wong FHC, Liu T, Leung DKY, Zhang AY, Au WSH, Kwok WW, Shum AKY, Wong GHY, Lum TYS
Consuming Information Related to COVID-19 on Social Media Among Older Adults and Its Association With Anxiety, Social Trust in Information, and COVID-Safe Behaviors: Cross-sectional Telephone Survey
J Med Internet Res 2021;23(2):e26570
URL: http://www.jmir.org/2021/2/e26570/
doi:10.2196/26570
PMID:33523825

Original Paper

Exposure to COVID-19-Related Information and its Association With Mental Health Problems in Thailand: Nationwide, Cross-sectional Survey Study

Pajaree Mongkhon[1,2], PhD; Chidchanok Ruengorn[2,3], PhD; Ratanaporn Awiphan[2,3], PhD; Kednapa Thavorn[2,4,5,6], PhD; Brian Hutton[4,5,6], PhD; Nahathai Wongpakaran[7], MD; Tinakon Wongpakaran[7], MD; Surapon Nochaiwong[2,3], PharmD

[1]Division of Pharmacy Practice, Department of Pharmaceutical Care; Unit of Excellence on Research in Health Outcomes and Patient Safety in Elderly, School of Pharmaceutical Sciences, University of Phayao, Phayao, Thailand

[2]Pharmacoepidemiology and Statistics Research Center, Faculty of Pharmacy, Chiang Mai University, Chiang Mai, Thailand

[3]Department of Pharmaceutical Care, Faculty of Pharmacy, Chiang Mai University, Chiang Mai, Thailand

[4]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

[5]Institute of Clinical and Evaluative Sciences, ICES uOttawa, Ottawa, ON, Canada

[6]Ottawa Hospital Research Institute, Ottawa Hospital, Ottawa, ON, Canada

[7]Department of Psychiatry, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

Corresponding Author:
Surapon Nochaiwong, PharmD
Department of Pharmaceutical Care
Faculty of Pharmacy
Chiang Mai University
239, Suthep Road
Chiang Mai, 50200
Thailand
Phone: 66 899973365
Email: surapon.nochaiwong@gmail.com

Abstract

Background: The COVID-19 pandemic has had a negative impact on both the physical and mental health of individuals worldwide. Evidence regarding the association between mental health problems and information exposure among Thai citizens during the COVID-19 outbreak is limited.

Objective: This study aimed to explore the relationship between information exposure and mental health problems during the COVID-19 pandemic in Thailand.

Methods: Between April 21 and May 4, 2020, we conducted a cross-sectional, nationwide online survey of the general population in Thailand. We categorized the duration of exposure to COVID-19-related information as follows: <1 h/day (reference group), 1-2 h/day, and ≥3 h/day. Mental health outcomes were assessed using the Patient Health Questionnaire-9, the Generalized Anxiety Disorder-7 scale, the Perceived Stress Scale-10, and the Insomnia Severity Index for symptoms of depression, anxiety, perceived stress, and insomnia, respectively. Multivariable logistic regression models were used to evaluate the relationship between information exposure and the risk of developing the aforementioned symptoms. An ancillary analysis using multivariable multinomial logistic regression models was also conducted to assess the possible dose-response relationship across the severity strata of mental health problems.

Results: Of the 4322 eligible participants, 4004 (92.6%) completed the online survey. Of them, 1481 (37.0%), 1644 (41.1%), and 879 (22.0%) participants were exposed to COVID-19-related information for less than 1 hour per day, 1 to 2 hours per day, or 3 or more hours per day, respectively. The major source of information related to the COVID-19 pandemic was social media (95.3%), followed by traditional media (68.7%) and family members (34.9%). Those exposed to information for 3 or more hours per day had a higher risk of developing symptoms of depression (adjusted odds ratio [OR] 1.35, 95% CI 1.03-1.76; *P*=.03), anxiety (adjusted OR 1.88, 95% CI 1.43-2.46; *P*<.001), and insomnia (adjusted OR 1.52, 95% CI 1.17-1.97; *P*=.001) than people exposed to information for less than 1 hour per day. Meanwhile, people exposed to information for 1 to 2 hours per day were only at risk of developing symptoms of anxiety (adjusted OR 1.35, 95% CI 1.08-1.69; *P*=.008). However, no association was

found between information exposure and the risk of perceived stress. In the ancillary analysis, a dose-response relationship was observed between information exposure of 3 or more hours per day and the severity of mental health problems.

Conclusions: These findings suggest that social media is the main source of COVID-19-related information. Moreover, people who are exposed to information for 3 or more hours per day are more likely to develop psychological problems, including depression, anxiety, and insomnia. Longitudinal studies investigating the long-term effects of COVID-19-related information exposure on mental health are warranted.

(J Med Internet Res 2021;23(2):e25363) doi:10.2196/25363

KEYWORDS

coronavirus; COVID-19; insomnia; mental health; social media; depression; anxiety; stress; psychosocial problem

Introduction

On December 31, 2019, the world witnessed the occurrence of a new public health emergency, the COVID-19 outbreak, in Wuhan, China [1]. On March 11, 2020, the World Health Organization (WHO) declared COVID-19 a pandemic due to the rapid global spread of the causative virus [2]. As of January 6, 2021, approximately 86 million confirmed cases and over 1.8 million deaths due to COVID-19 were reported worldwide. In addition to the physical effects of COVID-19, the COVID-19 pandemic has negatively affected the mental health of the public globally [3-5].

During the outbreak, people may need some information from the media to better understand the situation and determine strategies to protect their health. Information-seeking behavior may reduce anxiety caused by uncertainty during a disease outbreak or disaster [6]. In contrast, excessive consumption of information provided by the media may create new problems. Large volumes of information may amplify the perception of risk, and consumption of fear-related information may have a negative impact on consumers who cannot discern real news from fake news or cannot obtain a more balanced view of the media coverage of said event [7]. This infodemic has the potential to affect the population's mental health and well-being. Many previous studies have illustrated that media exposure is associated with adverse psychological outcomes in different contexts, including bioterrorism [8], war [9], natural disasters [10], and mental health of the general population [11].

Since the start of the COVID-19 pandemic, people have been highly dependent on information from the media, especially those who are not directly affected by the disease. People who are quarantined or isolated may also experience psychological problems due to the widespread media coverage related to the COVID-19 outbreak as well as financial difficulties. In addition, patients with confirmed or suspected COVID-19 may experience disease progression and transmit the virus to their families and friends. Numerous studies have investigated the association between COVID-19-related information exposure and mental health [12-17]. A previous study conducted in Wuhan demonstrated that social media exposure was positively associated with anxiety and depression during the COVID-19 outbreak, after controlling for covariates [4]. These results were supported by those of studies conducted in Germany [18], Saudi Arabia [19], and China [20] indicating that the frequency and duration of media or information exposure may predispose individuals to mental distress. These findings highlight the need to address mental health problems as part of public health policy.

In Thailand, the government announced a lockdown on March 26, 2020, in an effort to stop the spread of COVID-19. This measure seemed to have greatly prevented or slowed down the nationwide spread of the disease. However, during this lockdown, people who were quarantined or isolated might have developed psychological distress and other mental health problems due to media information overload and fear of the effects of COVID-19. To our knowledge, no study has reported the relationship between information exposure and mental health in the general population in Thailand during the COVID-19 outbreak. Therefore, we conducted a cross-sectional nationwide online survey to investigate the relationship between information exposure and symptoms of depression, anxiety, stress, and insomnia during the COVID-19 pandemic in Thailand.

Methods

Study Design and Study Population

The *Health Outcomes and Mental Health Care Evaluation Survey: Under the Pandemic Situation of COVID-19 (HOME-COVID-19)* was a cross-sectional online survey administered via the SurveyMonkey platform. The HOME-COVID-19 study was specifically developed by our research group and comprised baseline sociodemographic characteristics and a set of measurement tools for evaluating mental health and psychosocial problems; we used the Thai versions of the validated measurement tools or tools developed by our team. The questionnaire was initially revised by a panel of health care professionals, including three epidemiologists, two psychiatrists, one social scientist, and two hospital directors. This was further validated by a pilot survey of 30 health care professionals and 30 individuals from the general population. Further details about the HOME-COVID-19 study have been provided in a previous study [21]. This was a nationwide survey conducted between April 21 and May 4, 2020. This survey was performed by first selecting a sample of individuals from the general population in Thailand. Convenience sampling and a snowball strategy were used for participant recruitment through various social media networks (ie, public websites, Facebook, LINE, Twitter, and Instagram). The characteristics of the target population and the questionnaire used were subsequently presented [21]. We included Thai citizens, permanent residents, and nonresidents with employment or work permits who could read and communicate in the Thai language and were at least

18 years of age during the survey period. We excluded individuals who lacked internet access, were unable to complete the online survey, and spent less than 2 minutes, or more than 60 minutes, completing the survey, which likely made their data invalid. Based on the HOME-COVID-19 survey protocol [21], the participants spent approximately 20 to 30 minutes completing the survey.

Written consent was obtained from the participants in the first section of the online survey before completing the questionnaire. This study was approved by the Research Ethics Committees of the Faculties of Public Health (ET010/2020) and Pharmacy (23/2563), Chiang Mai University, Thailand. The study was reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology statement [22] and the Checklist for Reporting Results of Internet E-Surveys guidelines [23].

Measurements

COVID-19-Related Information Exposure

Information on the duration of information exposure was obtained by asking the participants how often they were exposed to news and information about the COVID-19 pandemic. The duration of information exposure was categorized as follows: <1 h/day, 1-2 h/day, or ≥3 h/day. Sources of information were classified as traditional media (eg, newspaper, television, radio, etc), social media (eg, Facebook, Twitter, blogs, etc), colleagues and neighbors, family members, government organizations (eg, Ministry of Health), and others.

Mental Health Outcomes

The mental health outcomes of interest included symptoms of depression, anxiety, perceived stress, and insomnia during the COVID-19 pandemic. The aforementioned symptoms were assessed using the Thai versions of the validated measurement tools. The measurement tools used in this study were as follows:

1. The Patient Health Questionnaire-9 (PHQ-9) consists of nine items that reflect the severity of depression symptoms. The score ranges from 0 to 27, and an overall PHQ-9 score of 9 or higher indicates depression (see Multimedia Appendix 1) [24].
2. The Generalized Anxiety Disorder-7 (GAD-7) scale comprises seven items that reflect the severity of anxiety symptoms. The total score ranges from 0 to 21, and a GAD-7 score of 5 or higher indicates anxiety (see Multimedia Appendix 2) [25].
3. The Perceived Stress Scale-10 (PSS-10) is a 5-point Likert scale consisting of 10 items. The scores range from 0 to 40, and a score of 14 or higher indicates perceived stress (see Multimedia Appendix 3) [26].
4. The Insomnia Severity Index (ISI) is an 8-item scale used to assess the nature, severity, and impact of insomnia. The scores range from 0 to 28 points and vary based on the severity of insomnia. In our study, an ISI score of 7 or higher indicated insomnia symptoms (see Multimedia Appendix 4) [27].

Power

Using the HOME-COVID-19 survey protocol, the target sample size for this study was estimated based on the findings of previous studies conducted during the COVID-19 outbreak, which reported a prevalence rate of mental health and psychosocial problems of 3.3% to 75.5% [21]. To compensate for a design effect of 2.0 and an estimated response rate of 60%, a minimum sample size of 2492 participants was obtained to achieve a power of 80% and a type I error α level of .05. However, there was no restriction on the maximum number of participants in the online survey.

Statistical Analysis

Descriptive analyses were performed to describe the characteristics of participants among the three information exposure groups. Categorical variables were expressed as the number (%), while continuous variables were expressed as the mean (SD) or median (IQR), as appropriate. Intergroup differences in the duration of COVID-19-related information exposure were tested using the χ^2 or Fisher exact tests; analysis of covariance and the Kruskal-Wallis test were used to evaluate categorical and continuous variables, respectively.

For the primary analysis, multivariable logistic regression models were used to explain the association between the duration of daily information exposure and the risk of depression (PHQ-9 score 9), anxiety (GAD-7 score 5), perceived stress (PSS-10 score 14), and insomnia (ISI score 8) during the COVID-19 outbreak, after controlling for the covariates of each outcome. The covariates included age, sex, marital status, education level, religion, occupation, region of residence, living status, reimbursement scheme, mental illness history, chronic noncommunicable disease history, income loss, financial problems, confirmed cases in the community, working from home, quarantine status, fear of COVID-19, and resilient coping. The degree of fear against COVID-19 was measured using a 10-point rating scale. The scores were then grouped as follows: no or minimal fear (0-3 points), moderate fear (4-6 points), and severe fear (7-10 points). Resilient coping was defined using the Brief Resilient Coping Scale: low-resilient copers (4-13 points), moderate-resilient copers (14-16 points), and high-resilient copers (17-20 points) [28].

An auxiliary analysis was conducted using multivariable multinomial logistic regression models to explain the association between the duration of information exposure and the severity of mental health problems. Sensitivity analysis was performed using a multivariable linear regression model to confirm the linear relationship between the duration of COVID-19-related information exposure and the risk of mental health problems. Meanwhile, a multivariable ordinal logistic regression model was used to examine the association between the duration of information exposure and severity strata of each mental health problem while controlling for the aforementioned covariates.

The effect estimates were presented as odds ratios (ORs), along with their corresponding 95% CIs, and weighted to match the estimates for the national population and internet users in all models based on data from the National Statistical Office under the Ministry of Information and Communication Technology.

All data were analyzed using Stata, version 14.0 (StataCorp LP). A 2-tailed test with $P<.05$ was considered statistically significant.

Results

Baseline Characteristics

A total of 4997 participants were invited to complete an online survey, but only 4381 participants responded to the survey. Among 4381 participants, 59 individuals who were under 18 years or age at the time of the survey and/or spent less than 2 minutes, or more than 60 minutes, completing the survey were excluded. Of the 4322 participants who met the eligibility criteria, 4004 completed the online survey (see Figure 1). Of the 4004 participants who completed the survey, 2619 (65.4%) were female, 1231 (30.7%) were male, and 154 (3.8%) selected *other* for the sex question. The mean age was 29.1 (SD 10.8) years; moreover, 3208 (80.1%) participants were single. As shown in Table 1, 127 (3.2%) participants had completed junior high school, while 1893 (47.3%) had completed high vocational education. Among the 4004 participants who completed the survey, 1589 (39.7%) were college students, 526 (13.1%) were government or state enterprise employees, and 500 (12.5%) were private enterprise employees. The participants were mainly from the noncapital city and its environs (2579/4004, 64.4%) in Thailand, whereas 1425 (35.6%) were from the capital city and its environs. Only 383 (9.6%) participants had no fear of COVID-19, whereas 1940 (48.4%) had a severe fear of COVID-19. The majority of the participants were low-resilient copers (1756 /4004, 43.9%), while 678 (16.9%) were highly resilient.

Figure 1. Flow diagram for study participants. HOME-COVID-19: Health Outcomes and Mental Health Care Evaluation Survey: Under the Pandemic Situation of COVID-19.

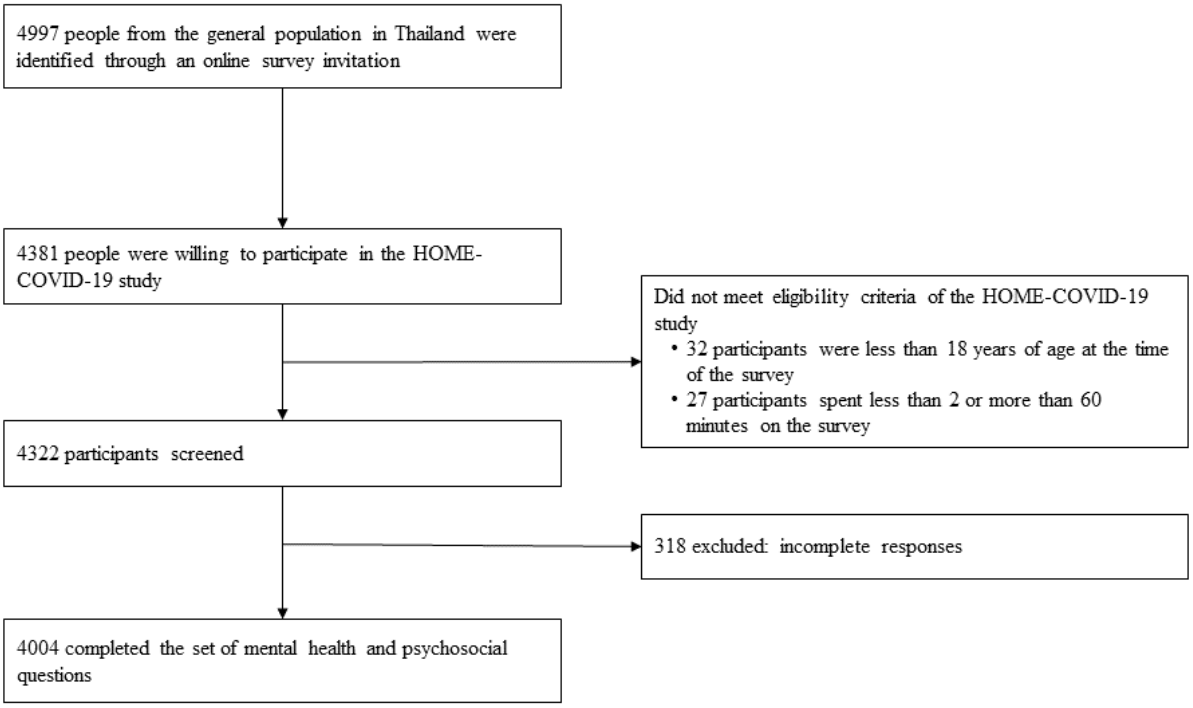


Table 1. Participant characteristics and the duration of exposure to COVID-19-related information in Thailand, from April 21 to May 4, 2020.

Participant characteristics	Overall (N=4004), n (%)	Duration of exposure to COVID-19-related information			
		<1 h/day (n=1481)	1-2 h/day (n=1644)	≥3 h/day (n=879)	*P* value
Sociodemographic data, n (%)					
Age					
Total sample (years), mean (SD)	29.1 (10.8)	27.8 (9.8)	28.9 (10.8)	31.5 (12.1)	<.001
≤30 years	2659 (66.4)	1028 (69.4)	1112 (67.6)	519 (59.0)	<.001
31-50 years	1088 (27.2)	392 (26.5)	427 (26.0)	269 (30.6)	
≥51 years	257 (6.4)	61 (4.1)	105 (6.4)	91 (10.4)	
Sex					
Male	1231 (30.7)	485 (32.8)	500 (30.4)	246 (28.0)	.12
Female	2619 (65.4)	936 (63.2)	1087 (66.1)	596 (67.8)	
Other	154 (3.9)	60 (4.0)	57 (3.5)	37 (4.2)	
Marital status					
Single	3208 (80.1)	1228 (82.9)	1335 (81.2)	645 (73.4)	<.001
Married or domestic partnership	693 (17.3)	219 (14.8)	277 (16.8)	197 (22.4)	
Divorced, widowed, or separated	103 (2.6)	34 (2.3)	32 (2.0)	37 (4.2)	
Education level					
Illiterate, primary school, or junior high school	127 (3.2)	54 (3.6)	46 (2.8)	27 (3.1)	<.001
Senior high school, diploma, or high vocational school	1893 (47.3)	799 (54.0)	749 (45.6)	345 (39.2)	
Bachelor's degree	1559 (38.9)	511 (34.5)	648 (39.4)	400 (45.5)	
Higher education	425 (10.6)	117 (7.9)	201 (12.2)	107 (12.2)	
Religion					
Irreligious	375 (9.4)	138 (9.3)	171 (10.4)	66 (7.5)	.23
Buddhist	3454 (86.3)	1271 (85.8)	1402 (85.3)	781 (88.9)	
Christian	100 (2.5)	43 (2.9)	42 (2.5)	15 (1.7)	
Muslim	70 (1.7)	26 (1.8)	28 (1.7)	16 (1.8)	
Other	5 (0.1)	3 (0.2)	1 (0.1)	1 (0.1)	
Occupation					
Unemployed or retired	391 (9.8)	130 (8.8)	148 (9.0)	113 (12.9)	<.001
Farmer or laborer	451 (11.3)	195 (13.2)	156 (9.5)	100 (11.4)	
Self-employed	415 (10.4)	146 (9.9)	154 (9.4)	115 (13.1)	
Government or state enterprise employee	526 (13.1)	139 (9.4)	260 (15.8)	127 (14.4)	
College student	1589 (39.7)	667 (45.0)	659 (40.1)	263 (29.9)	
Private enterprise employee	500 (12.5)	165 (11.1)	211 (12.8)	124 (14.1)	
Freelance or other	132 (3.3)	39 (2.6)	56 (3.4)	37 (4.2)	
Region of residence					
Capital city and its environs	1425 (35.6)	496 (33.5)	623 (37.9)	306 (34.8)	.03
Noncapital city and its environs	2579 (64.4)	985 (66.5)	1021 (62.1)	573 (65.2)	
Living status					
Alone	576 (14.4)	195 (13.2)	239 (14.5)	142 (16.2)	.25
With family	3164 (79.0)	1179 (79.6)	1299 (79.0)	686 (78.0)	
With others	264 (6.6)	107 (7.2)	106 (6.5)	51 (5.8)	

Participant characteristics	Overall (N=4004), n (%)	Duration of exposure to COVID-19-related information			
		<1 h/day (n=1481)	1-2 h/day (n=1644)	≥3 h/day (n=879)	*P* value
Reimbursement scheme					
Government or state enterprises	539 (13.5)	172 (11.6)	254 (15.5)	113 (12.9)	.008
Universal coverage scheme	1329 (33.2)	524 (35.4)	515 (31.3)	290 (33.0)	
Social security scheme	1161 (29.0)	423 (28.6)	460 (28.0)	278 (31.6)	
Self-payment or other	975 (24.3)	362 (24.4)	415 (25.2)	198 (22.5)	
History of mental illness					
Yes	359 (9.0)	125 (8.4)	142 (8.6)	92 (10.5)	.21
No	3645 (91.0)	1356 (91.6)	1502 (91.4)	787 (89.5)	
History of chronic noncommunicable disease[a]					
Yes	599 (15.0)	182 (12.3)	218 (13.3)	199 (22.6)	<.001
No	3405 (85.0)	1299 (87.7)	1426 (86.7)	680 (77.4)	
Economic burden and issues with regard to the COVID-19 outbreak, n (%)					
Income loss during the COVID-19 outbreak					
Yes	1664 (41.6)	524 (35.4)	660 (40.2)	480 (54.6)	<.001
No	2340 (58.4)	957 (64.6)	984 (59.8)	399 (45.4)	
Financial problems during the COVID-19 outbreak					
Yes	2012 (50.2)	668 (45.1)	800 (48.7)	544 (61.9)	<.001
No	1992 (49.8)	813 (54.9)	844 (51.3)	335 (38.1)	
Confirmed cases in the community					
No	2562 (64.0)	888 (60.0)	1067 (64.9)	607 (69.1)	<.001
Yes	641 (16.0)	246 (16.6)	260 (15.8)	135 (15.4)	
Not known	801 (20.0)	347 (23.4)	317 (19.3)	137 (15.6)	
Working from home					
Yes	3139 (78.4)	1142 (77.1)	1298 (79.0)	699 (79.5)	.30
No	865 (21.6)	339 (22.9)	346 (21.0)	180 (20.5)	
Quarantine status					
Never	1781 (44.5)	610 (41.2)	751 (45.7)	420 (47.8)	.004
Past	1575 (39.3)	631 (42.6)	636 (38.7)	308 (35.0)	
Current	648 (16.2)	240 (16.2)	257 (15.6)	151 (17.2)	
Fear of COVID-19					
None or minimal	383 (9.6)	163 (11.0)	140 (8.5)	80 (9.1)	.04
Moderate	1681 (42.0)	593 (40.0)	729 (44.3)	359 (40.8)	
Severe	1940 (48.4)	725 (49.0)	775 (47.1)	440 (50.1)	
Resilient coping					
Low-resilient coper	1756 (43.9)	683 (46.1)	701 (42.6)	372 (42.3)	.12
Medium-resilient coper	1570 (39.2)	559 (37.7)	670 (40.8)	341 (38.8)	
High-resilient coper	678 (16.9)	239 (16.1)	273 (16.6)	166 (18.9)	

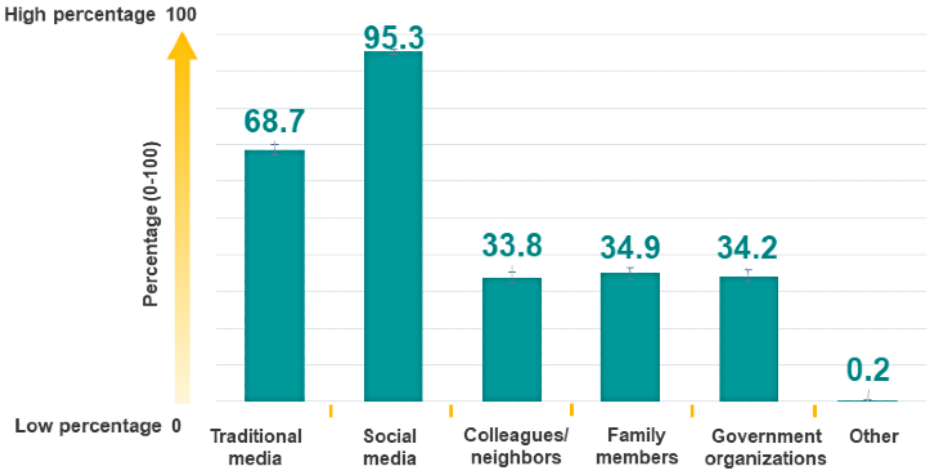
[a]Includes diabetes mellitus, hypertension, dyslipidemia, stroke and heart disease, chronic kidney disease, chronic lung disease, and cancer.

Information Exposure

Of the 4004 participants, 1644 (41.1%) were exposed to information 1 to 2 hours per day, followed by less than 1 hour per day (1481/4004, 37.0%) and 3 or more hours per day (879/4004, 22.0%). The sources of COVID-19-related information were social media (95.3%, 95% CI 94.6-95.9),

followed by traditional media (68.7%, 95% CI 67.2-70.1), and family members (34.9%, 95% CI 33.4-36.4) (see Figure 2).

Figure 2. Sources of COVID-19-related information in Thailand. One person can be exposed to more than one source.



Mental Health and Information Exposure

Table 2 presents the results of the multivariable logistic regression analysis that was performed to explain the association between the duration of information exposure and the risk of mental health problems during the COVID-19 pandemic, after controlling for covariates. In the primary analysis, participants who were exposed to COVID-19-related information for 3 or more hours per day had higher odds of developing symptoms of depression than those exposed to COVID-19-related information for less than 1 hour per day (adjusted OR 1.35, 95% CI 1.03-1.76; *P*=.03). Similarly, participants who were exposed to COVID-19-related information for 3 or more hours per day were more likely to develop anxiety and insomnia than those exposed to COVID-19-related information for less than 1 hour per day (adjusted OR 1.88, 95% CI 1.43-2.46; *P*<.001; and adjusted OR 1.52, 95% CI 1.17-1.97; *P*=.001, respectively). Meanwhile, participants who were exposed to COVID-19-related information for 1 to 2 hours per day only had a risk of developing anxiety symptoms (adjusted OR 1.35, 95% CI 1.08-1.69; *P*=.008). No association was observed between the duration of information exposure and the risk of perceived stress in the general population during the study period in Thailand.

Table 2. Mental health outcomes according to the duration of exposure to COVID-19-related information: multivariable logistic regression model.

Symptoms developed and duration of exposure to COVID-19-related information	Number of cases, n (%)	Adjusted OR[a] (95% CI)[b]	*P* value
Depression: Patient Health Questionnaire-9			
<1 h/day (n=1481)	567 (38.3)	Reference (1.00)	N/A[c]
1-2 h/day (n=1644)	696 (42.3)	1.08 (0.86-1.35)	.52
≥3 h/day (n=879)	391 (44.5)	1.35 (1.03-1.76)	.03
Anxiety: Generalized Anxiety Disorder-7			
<1 h/day (n=1481)	578 (39.0)	Reference (1.00)	N/A
1-2 h/day (n=1644)	701 (42.6)	1.35 (1.08-1.69)	.008
≥3 h/day (n=879)	446 (50.7)	1.88 (1.43-2.46)	<.001
Stress: Perceived Stress Scale-10			
<1 h/day (n=1481)	1076 (72.7)	Reference (1.00)	N/A
1-2 h/day (n=1644)	1200 (73.0)	0.96 (0.74-1.23)	.72
≥3 h/day (n=879)	640 (72.8)	1.01 (0.75-1.36)	.93
Insomnia: Insomnia Severity Index			
<1 h/day (n=1481)	760 (51.3)	Reference (1.00)	
1-2 h/day (n=1644)	877 (53.3)	1.04 (0.84-1.29)	.71
≥3 h/day (n=879)	520 (59.2)	1.52 (1.17-1.97)	.001

[a]OR: odds ratio.

[b]ORs and corresponding 95% CIs are presented as weighted according to the national population and internet users in Thailand. ORs were adjusted for age, sex, marital status, education level, religion, occupation, region of residence, living status, reimbursement scheme, history of mental illness, history of chronic noncommunicable disease, income loss, financial problems, confirmed cases in the community, working from home, quarantine status, fear of COVID-19, and resilient coping.

[c]N/A: not applicable.

In an ancillary analysis, using a multivariable multinomial logistic regression model (<1 h/day of information exposure [reference group]), a dose-response relationship was noted between exposure to COVID-19-related information for 3 or more hours per day and the severity of mental health problems, particularly anxiety symptoms (see Table 3). For instance, individuals exposed to COVID-19-related information for 3 or more hours per day demonstrated substantial effect modification in terms of the severity strata of anxiety symptoms: the adjusted ORs were 1.35 (95% CI 0.99-1.83; *P*=.06), 2.87 (95% CI 1.91-4.33; *P*<.001), and 4.45 (95% CI 2.45-8.08; *P*<.001) for mild, moderate, and severe anxiety symptoms, respectively.

Table 3. Multivariable multinomial logistic regression model results of the duration of exposure to COVID-19-related information and severity of mental health problems.

Severity of mental health problems	Number of cases (N=4004), n (%)	Duration of exposure to COVID-19-related information			
		1-2 h/day		≥3 h/day	
		Estimated adjusted OR[a] (95% CI)[b]	*P* value	Estimated adjusted OR (95% CI)[b]	*P* value
Depression: Patient Health Questionnaire-9					
None or minimal (<9 points)	2350 (58.7)	Reference (1.00)	N/A[c]	Reference (1.00)	N/A
Mild (9-14 points)	1024 (25.6)	1.05 (0.82-1.34)	.70	1.18 (0.86-1.60)	.31
Moderate (15-19 points)	389 (9.7)	1.10 (0.76-1.60)	.60	1.49 (0.97-2.28)	.07
Severe (≥20 points)	241 (6.0)	1.13 (0.70-1.84)	.61	2.23 (1.27-3.92)	.005
Anxiety: Generalized Anxiety Disorder-7					
None or minimal (<5 points)	2279 (56.9)	Reference (1.00)	N/A	Reference (1.00)	N/A
Mild (5-9 points)	1117 (27.9)	1.31 (1.03-1.66)	.03	1.35 (0.99-1.83)	.06
Moderate (10-14 points)	411 (10.3)	1.33 (0.91-1.93)	.14	2.87 (1.91-4.33)	<.001
Severe (≥15 points)	197 (4.9)	1.79 (1.06-3.02)	.03	4.45 (2.45-8.08)	<.001
Stress: Perceived Stress Scale-10					
None or low (<14 points)	1088 (27.2)	Reference (1.00)	N/A	Reference (1.00)	N/A
Moderate (14-26 points)	2550 (63.7)	0.93 (0.72-1.20)	.59	0.96 (0.71-1.30)	.80
High (≥27 points)	366 (9.1)	1.22 (0.80-1.88)	.36	1.76 (1.09-2.85)	.02
Insomnia: Insomnia Severity Index					
None or minimal (<8 points)	1847 (46.1)	Reference (1.00)	N/A	Reference (1.00)	N/A
Mild (8-14 points)	1576 (39.4)	0.99 (0.79-1.24)	.94	1.41 (1.07-1.86)	.02
Moderate (15-21 points)	486 (12.1)	1.29 (0.91-1.83)	.15	2.03 (1.34-3.07)	.001
Severe (≥22 points)	95 (2.4)	0.89 (0.45-1.76)	.74	1.48 (0.69-3.18)	.31

[a]OR: odds ratio.

[b]Multinomial logistic regression was performed using information exposure of <1 h/day as the reference group. ORs and corresponding 95% CIs are presented as weighted according to the national population and internet users in Thailand. ORs were adjusted for age, sex, marital status, education level, religion, occupation, region of residence, living status, reimbursement scheme, history of mental illness, history of chronic noncommunicable disease, income loss, financial problems, confirmed cases in the community, working from home, quarantine status, fear of COVID-19, and resilient coping.

[c]N/A: not applicable.

Sensitivity Analysis

Sensitivity analysis was performed by considering the mental health outcomes. A linear relationship was seen between adverse mental health outcomes and the duration of COVID-19-related information exposure, particularly for 3 or more hours per day (see Multimedia Appendix 5, Table S1). Moreover, after using a multivariable ordinal logistic regression model, our findings remained the same compared to those in the primary analysis (see Multimedia Appendix 5, Table S2).

Data Sharing

Data will be shared upon reasonable request and with permission according to the Health Outcomes and Mental Health Care Evaluation Survey Research Group data release policy.

Discussion

Principal Findings

This nationwide online survey was conducted in Thailand. The results indicated that many Thai people used social media as the main source of information during the COVID-19 pandemic. Most participants were exposed to COVID-19-related information for 1 to 2 hours per day. Participants who were exposed to information for 3 or more hours per day had a higher risk of developing depression, anxiety, and insomnia.

Our results were in line with those of a previous study conducted in Wuhan, China, demonstrating that social media exposure was associated with higher odds of developing anxiety and depression [17]. Moreover, previous studies [17,18,20,29] showed that not only the frequency but also the daily duration of exposure to COVID-19-related information was associated

with psychosocial problems [29], which corroborated with our study results. The results of a survey conducted in Germany indicated a positive correlation between COVID-19-related media exposure and the severity of nonspecific anxiety and depression [18]. Similar results were reported in a previous study conducted among the Saudi Arabian general population [19].

In essence, during a lockdown or social isolation, people are exposed to a lot of pandemic-related information, especially through social media, which predisposes them to mental health problems. During a severe social disruption, the mass media are expected to satisfy the needs of individuals for information about operationalized guidance to the public, the response of organizations, and an exchange of views with others [30-32]. However, the amount of information that circulated during the COVID-19 pandemic exceeded the information demand. A recent study indicated that repeated media exposure could lead to anxiety due to the effect of vicarious traumatization [31]. Furthermore, media exposure can cause an infodemic, which is defined as fake news, misinformation, and conspiracy theories, making it difficult for individuals to find trusted information [33]. This infodemic may negatively impact an individual's mental health.

In Thailand, the government issued a national emergency decree, which was put into effect beginning on March 26, 2020, and announced a nationwide curfew on April 3, 2020. People were requested to stay indoors and limit all social contacts. The government also requested that people should wear face masks, practice social distancing, and remain indoors from 10 PM to 4 AM [34]. Due to the implementation of curfews and isolation measures, the number of persons with suspected or confirmed COVID-19 decreased from mid-January to April 6, 2020, which indicated a positive impact of the lockdown on viral transmission reduction and epidemic control [35]. However, the quarantine and isolation measures, coupled with media coverage exposure, increased anxiety and fear in the people, thereby affecting their mental health. Moreover, during the national lockdown and home confinement, the community movement in Thailand had eventually decreased, which resulted in home-based exposure of people to huge amounts of COVID-19-related media information.

Therefore, during a public health crisis, several actions should be taken regarding crisis-related information. First, the media should convey information to the public without sensationalizing the situation and without providing disturbing images to help prevent mental distress. Second, the public should rely on trustworthy sources of information, such as the Centers for Disease Control and Prevention or the WHO, for accurate information. Third, health care providers should play an important role in informing people about practicing protective behaviors [36]. Fourth, public awareness campaigns should be conducted, focusing on the maintenance of mental health in the prevailing situation [37]. Finally, people should spend time with family members who engage in different healthy exercises and sports activities, follow a schedule and routine in daily life, and reduce their time spent on traditional and social media to remain healthy [37]. In terms of mental health policies during the COVID-19 pandemic, the United Nations suggested the application of a whole-of-society approach to promote, protect, and care for an individual's mental health [38]. Furthermore, public policy solutions would ensure the widespread availability of emergency mental health and psychosocial support. All affected communities require quality mental health services to help their society recover from the effects of the COVID-19 pandemic.

Strengths and Limitations

This study had several strengths. To our knowledge, this is the first nationwide survey on mental health based on information exposure during the COVID-19 outbreak in Thailand. From a methodological point of view, the analyses were also performed using rigorous statistical approaches to confirm the main findings.

However, our study has some limitations. First, this was a cross-sectional study; therefore, a causal relationship could not be accurately drawn. Second, the survey was conducted online, which is appropriate for short-term evaluation; therefore, the results reflect short-term relationships. Hence, additional longitudinal studies investigating the long-term relationship between mental health and COVID-19-related information exposure are warranted. Third, the survey was conducted over the internet, which is subject to bias due to nonresponse effects and selection bias. Moreover, as information exposure was self-reported, recall bias may be present, and the measures may be influenced by social desirability. Fourth, not everyone in Thailand can access the internet and there could be significant demographic differences between those who can access the internet and those who have limited access to the internet. Thus, selection bias or limited representativeness of the Thai population may be present. Fifth, the majority of the participants were young adults. Only a few older adults participated in the survey; thus, some respondent bias could not be excluded. In addition, the majority of participants (66.4%) were below 30 years of age, which might not reflect the Thai society in its entirety. However, internet usage data in Thailand have demonstrated that workers spend the highest average number of hours surfing the internet per day and account for the highest proportion of internet users [39]. The internet usage data of this group of people may potentially represent the internet consumption of the Thai population. Finally, we lacked information on risk perception and self-care behaviors. This information could help encourage self-improvement and self-management during the COVID-19 pandemic, which may improve an individual's overall mental health and wellness.

Conclusions

Our findings from the nationwide online survey indicated a positive association between information exposure during the COVID-19 pandemic and the occurrence of symptoms of depression, anxiety, and insomnia. The strength of the association increased with the duration of media exposure. However, prospective longitudinal studies are needed to investigate the long-term relationship between information exposure and mental health during and after the COVID-19 pandemic.

Acknowledgments

This work was partially supported by the Chiang Mai University, Thailand, with support provided to SN. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of Chiang Mai University. No funding source had any role in the study concept and design; data collection, analysis, and interpretation; writing of the report; or in the decision to submit for publication. We would like to thank all staff and research personnel from the Pharmacoepidemiology and Statistics Research Center, Chiang Mai University, Thailand. Particular thanks are given to the study participants.

Authors' Contributions

PM and SN had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. PM and SN were responsible for the study concept and design. PM, CR, and SN were responsible for the acquisition, analysis, or interpretation of data. PM and SN were responsible for drafting of the manuscript. Critical revision of the manuscript for important intellectual content was performed by KT, BH, NW, and TW. Statistical analysis was performed by PM, CR, and SN. Administrative, technical, or material support was provided by RA. Study supervision was provided by SN.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient Health Questionnaire-9 (PHQ-9).
[DOCX File , 17 KB - jmir_v23i2e25363_app1.docx]

Multimedia Appendix 2
Generalized Anxiety Disorder-7 (GAD-7) scale.
[DOCX File , 17 KB - jmir_v23i2e25363_app2.docx]

Multimedia Appendix 3
Perceived Stress Scale-10 (PSS-10).
[DOCX File , 18 KB - jmir_v23i2e25363_app3.docx]

Multimedia Appendix 4
Insomnia Severity Index (ISI).
[DOCX File , 18 KB - jmir_v23i2e25363_app4.docx]

Multimedia Appendix 5
Sensitivity analysis.
[DOCX File , 20 KB - jmir_v23i2e25363_app5.docx]

References

1. Novel Coronavirus (2019-nCoV): Situation Report - 1. Geneva, Switzerland: World Health Organization; 2020 Jan 21. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf [accessed 2020-09-27]
2. Coronavirus Disease 2019 (COVID-19): Situation Report – 51. Geneva, Switzerland: World Health Organization; 2020 Mar 11. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf [accessed 2020-09-27]
3. Bao Y, Sun Y, Meng S, Shi J, Lu L. 2019-nCoV epidemic: Address mental health care to empower society. Lancet 2020 Feb 22;395(10224):e37-e38 [FREE Full text] [doi: 10.1016/S0140-6736(20)30309-3] [Medline: 32043982]
4. Kang L, Li Y, Hu S, Chen M, Yang C, Yang BX, et al. The mental health of medical workers in Wuhan, China dealing with the 2019 novel coronavirus. Lancet Psychiatry 2020 Mar;7(3):e14. [doi: 10.1016/s2215-0366(20)30047-x]
5. Shigemura J, Ursano RJ, Morganstein JC, Kurosawa M, Benedek DM. Public responses to the novel 2019 coronavirus (2019-nCoV) in Japan: Mental health consequences and target populations. Psychiatry Clin Neurosci 2020 Apr;74(4):281-282 [FREE Full text] [doi: 10.1111/pcn.12988] [Medline: 32034840]
6. Heath RL, Gay CD. Risk communication. Manag Commun Q 2016 Aug 17;10(3):342-372. [doi: 10.1177/0893318997010003004]
7. Kasperson RE, Renn O, Slovic P, Brown HS, Emel J, Goble R, et al. The social amplification of risk: A conceptual framework. Risk Anal 1988 Jun;8(2):177-187 [FREE Full text] [doi: 10.1111/j.1539-6924.1988.tb01168.x]

8. Dougall AL, Hayward MC, Baum A. Media exposure to bioterrorism: Stress and the anthrax attacks. Psychiatry 2005;68(1):28-42. [doi: 10.1521/psyc.68.1.28.64188] [Medline: 15899708]
9. Bodas M, Siman-Tov M, Peleg K, Solomon Z. Anxiety-inducing media: The effect of constant news broadcasting on the well-being of Israeli television viewers. Psychiatry 2015;78(3):265-276. [doi: 10.1080/00332747.2015.1069658] [Medline: 26391834]
10. Hall BJ, Xiong YX, Yip PSY, Lao CK, Shi W, Sou EKL, et al. The association between disaster exposure and media use on post-traumatic stress disorder following Typhoon Hato in Macao, China. Eur J Psychotraumatol 2019;10(1):1558709 [FREE Full text] [doi: 10.1080/20008198.2018.1558709] [Medline: 30693078]
11. Kross E, Verduyn P, Demiralp E, Park J, Lee DS, Lin N, et al. Facebook use predicts declines in subjective well-being in young adults. PLoS One 2013;8(8):e69841 [FREE Full text] [doi: 10.1371/journal.pone.0069841] [Medline: 23967061]
12. Liu JCJ, Tong EMW. The relation between official WhatsApp-distributed COVID-19 news exposure and psychological symptoms: Cross-sectional survey study. J Med Internet Res 2020 Sep 25;22(9):e22142 [FREE Full text] [doi: 10.2196/22142] [Medline: 32877349]
13. Riehm KE, Holingue C, Kalb LG, Bennett D, Kapteyn A, Jiang Q, et al. Associations between media exposure and mental distress among US adults at the beginning of the COVID-19 pandemic. Am J Prev Med 2020 Nov;59(5):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2020.06.008] [Medline: 33011008]
14. Bazán P, Azevedo NR, Dias J, Salvatierra V, Sanches L, Lacerda S. COVID-19 information exposure in digital media and implications for employees in the health care sector: Findings from an online survey. Einstein (Sao Paulo) 2020 Nov 26;18:eAO6127. [doi: 10.31744/einstein_journal/2020ao6127]
15. Lee JJ, Kang K, Wang MP, Zhao SZ, Wong JYH, O'Connor S, et al. Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and preventive behaviors: Cross-sectional online study. J Med Internet Res 2020 Nov 13;22(11):e22205 [FREE Full text] [doi: 10.2196/22205] [Medline: 33048825]
16. Ruiz-Frutos C, Ortega-Moreno M, Dias A, Bernardes JM, García-Iglesias JJ, Gómez-Salgado J. Information on COVID-19 and psychological distress in a sample of non-health workers during the pandemic period. Int J Environ Res Public Health 2020 Sep 24;17(19):6982 [FREE Full text] [doi: 10.3390/ijerph17196982] [Medline: 32987712]
17. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
18. Bendau A, Petzold MB, Pyrkosch L, Mascarell Maricic L, Betzler F, Rogoll J, et al. Associations between COVID-19 related media consumption and symptoms of anxiety, depression and COVID-19 related fear in the general population in Germany. Eur Arch Psychiatry Clin Neurosci 2020 Jul 20:1-9 [FREE Full text] [doi: 10.1007/s00406-020-01171-6] [Medline: 32691135]
19. Zakout Y, Alreshidi F, Elsaid R, Ahmed H. The magnitude of COVID-19 related stress, anxiety and depression associated with intense mass media coverage in Saudi Arabia. AIMS Public Health 2020;7(3):664-678 [FREE Full text] [doi: 10.3934/publichealth.2020052] [Medline: 32968685]
20. Yao H. The more exposure to media information about COVID-19, the more distressed you will feel. Brain Behav Immun 2020 Jul;87:167-169 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.031] [Medline: 32413557]
21. Nochaiwong S, Ruengorn C, Awiphan R, Ruanta Y, Boonchieng W, Nanta S. Mental health circumstances among health care workers and general public under the pandemic situation of COVID-19 (HOME-COVID-19). Medicine 2020 Jun 26;99(26):e20751. [doi: 10.1097/md.0000000000020751]
22. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808 [FREE Full text] [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
23. Eysenbach G. Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
24. Lotrakul M, Sumrithe S, Saipanish R. Reliability and validity of the Thai version of the PHQ-9. BMC Psychiatry 2008 Jun 20;8(1):1. [doi: 10.1186/1471-244x-8-46]
25. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
26. Wongpakaran N, Wongpakaran T. The Thai version of the PSS-10: An investigation of its psychometric properties. Biopsychosoc Med 2010 Jun 12;4:6 [FREE Full text] [doi: 10.1186/1751-0759-4-6] [Medline: 20540784]
27. Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: Psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 01;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
28. Sinclair VG, Wallston KA. The development and psychometric evaluation of the Brief Resilient Coping Scale. Assessment 2004 Mar;11(1):94-101. [doi: 10.1177/1073191103258144] [Medline: 14994958]
29. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: A web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]

30. Ball-Rokeach S, DeFleur M. A dependency model of mass-media effects. Communic Res 2016 Jun 30;3(1):3-21. [doi: 10.1177/009365027600300101]
31. Liu C, Liu Y. Media exposure and anxiety during COVID-19: The mediation effect of media vicarious traumatization. Int J Environ Res Public Health 2020 Jun 30;17(13):4720 [FREE Full text] [doi: 10.3390/ijerph17134720] [Medline: 32630054]
32. Lowrey W. Media dependency during a large-scale social disruption: The case of September 11. Mass Commun Soc 2004 Jul;7(3):339-357. [doi: 10.1207/s15327825mcs0703_5]
33. The Lancet Infectious Diseases. The COVID-19 infodemic. Lancet Infect Dis 2020 Aug;20(8):875 [FREE Full text] [doi: 10.1016/S1473-3099(20)30565-X] [Medline: 32687807]
34. Coronavirus Disease 2019 (COVID-19): WHO Thailand Situation Report. Geneva, Switzerland: World Health Organization; 2020 Mar 25. URL: https://www.who.int/docs/default-source/searo/thailand/2020-03-25-tha-sitrep-32-covid19-final.pdf?sfvrsn=adce58ef_0 [accessed 2020-09-27]
35. Dechsupa S, Assawakosri S, Phakham S, Honsawek S. Positive impact of lockdown on COVID-19 outbreak in Thailand. Travel Med Infect Dis 2020;36:101802 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101802] [Medline: 32569811]
36. Garfin DR, Silver RC, Holman EA. The novel coronavirus (COVID-2019) outbreak: Amplification of public health consequences by media exposure. Health Psychol 2020 May;39(5):355-357 [FREE Full text] [doi: 10.1037/hea0000875] [Medline: 32202824]
37. Javed B, Sarwer A, Soto EB, Mashwani Z. The coronavirus (COVID-19) pandemic's impact on mental health. Int J Health Plann Manage 2020 Sep;35(5):993-996 [FREE Full text] [doi: 10.1002/hpm.3008] [Medline: 32567725]
38. Policy Brief: COVID-19 and the Need for Action on Mental Health. New York, NY: United Nations; 2020 May 13. URL: https://unsdg.un.org/sites/default/files/2020-05/UN-Policy-Brief-COVID-19-and-mental-health.pdf [accessed 2020-10-24]
39. Office of Policy and Strategic Affairs, Electronic Transactions Development Agency (ETDA), Ministry of Digital Economy and Society. Thailand Internet User Behavior 2019. Bangkok, Thailand: Electronic Transactions Development Agency (ETDA); 2020. URL: https://www.slideshare.net/ETDAofficialRegist/thailand-internet-user-behavior-2019 [accessed 2021-02-03]

Abbreviations

GAD-7: Generalized Anxiety Disorder-7
HOME-COVID-19: Health Outcomes and Mental Health Care Evaluation Survey: Under the Pandemic Situation of COVID-19
ISI: Insomnia Severity Index
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9
PSS-10: Perceived Stress Scale-10
WHO: World Health Organization

Edited by G Eysenbach; submitted 29.10.20; peer-reviewed by J Hoogervorst, A Bregenzer; comments to author 23.12.20; revised version received 12.01.21; accepted 29.01.21; published 12.02.21.

Please cite as:
Mongkhon P, Ruengorn C, Awiphan R, Thavorn K, Hutton B, Wongpakaran N, Wongpakaran T, Nochaiwong S
Exposure to COVID-19-Related Information and its Association With Mental Health Problems in Thailand: Nationwide, Cross-sectional Survey Study
J Med Internet Res 2021;23(2):e25363
URL: http://www.jmir.org/2021/2/e25363/
doi:10.2196/25363
PMID:33523828

Original Paper

Quantifying the Influence of Delay in Opinion Transmission of COVID-19 Information Propagation: Modeling Study

Fulian Yin[1], PhD; Xueying Shao[1], MD; Meiqi Ji[1], MD; Jianhong Wu[2], PhD

[1]College of Information and Communication Engineering, Communication University of China, Beijing, China

[2]Fields-CQAM Laboratory of Mathematics for Public Health, Laboratory for Industrial and Applied Mathematics, York University, Toronto, ON, Canada

Corresponding Author:
Jianhong Wu, PhD
Fields-CQAM Laboratory of Mathematics for Public Health
Laboratory for Industrial and Applied Mathematics
York University
4700 Keele St
Toronto, ON, M3J1P3
Canada
Phone: 1 416 736 5243
Email: wujhhida@hotmail.com

Abstract

Background: In a fast-evolving public health crisis such as the COVID-19 pandemic, multiple pieces of relevant information can be posted sequentially on a social media platform. The interval between subsequent posting times may have a different impact on the transmission and cross-propagation of the old and new information that results in a different peak value and a final size of forwarding users of the new information, depending on the content correlation and whether the new information is posted during the outbreak or quasi–steady-state phase of the old information.

Objective: This study aims to help in designing effective communication strategies to ensure information is delivered to the maximal number of users.

Methods: We developed and analyzed two classes of susceptible-forwarding-immune information propagation models with delay in transmission to describe the cross-propagation process of relevant information. A total of 28,661 retweets of typical information were posted frequently by each opinion leader related to COVID-19 with high influence (data acquisition up to February 19, 2020). The information was processed into discrete points with a frequency of 10 minutes, and the real data were fitted by the model numerical simulation. Furthermore, the influence of parameters on information dissemination and the design of a publishing strategy were analyzed.

Results: The current epidemic outbreak situation, epidemic prevention, and other related authoritative information cannot be timely and effectively browsed by the public. The ingenious use of information release intervals can effectively enhance the interaction between information and realize the effective diffusion of information. We parameterized our models using real data from Sina Microblog and used the parameterized models to define and evaluate mutual attractiveness indexes, and we used these indexes and parameter sensitivity analyses to inform optimal strategies for new information to be effectively propagated in the microblog. The results of the parameter analysis showed that using different attractiveness indexes as the key parameters can control the information transmission with different release intervals, so it is considered as a key link in the design of an information communication strategy. At the same time, the dynamic process of information was analyzed through index evaluation.

Conclusions: Our model can carry out an accurate numerical simulation of information at different release intervals and achieve a dynamic evaluation of information transmission by constructing an indicator system so as to provide theoretical support and strategic suggestions for government decision making. This study optimizes information posting strategies to maximize communication efforts for delivering key public health messages to the public for better outcomes of public health emergency management.

(J Med Internet Res 2021;23(2):e25734) doi:10.2196/25734

KEYWORDS

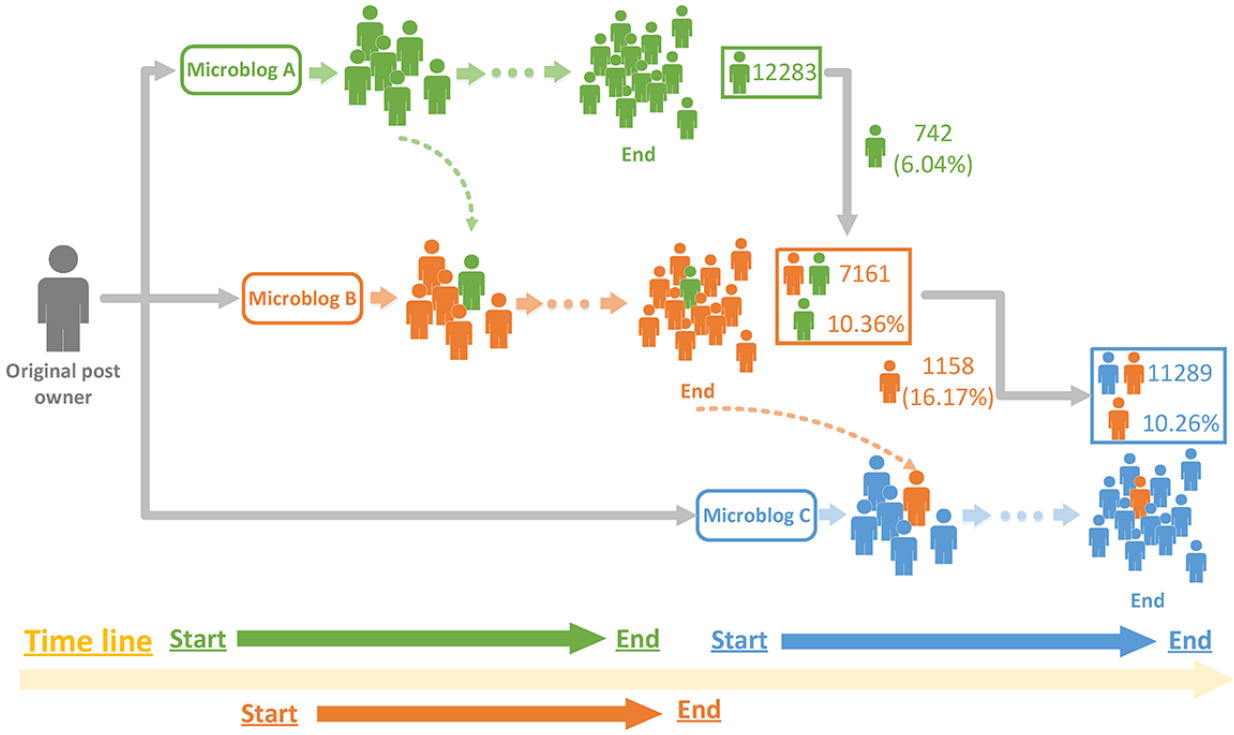
COVID-19; delay transmission; dynamic model; Sina Microblog; social media; communication; online health information; health information; public health; opinion; strategy; model; information transmission; delay; infodemiology; infoveillance

Introduction

In the absence of effective treatment or a vaccine, the success or failure of mitigating COVID-19 transmission in the population relies heavily on the effectiveness of social distancing, self-protection, case detection, quarantine, isolation, and testing. The effectiveness of these nonpharmaceutical interventions depends on the active participation and engagement of residents in the community, which is substantially influenced by the public opinion. Given the short disease transmission doubling time, the timing (and hence time lags) in the communication of critical public health information to the community has a profound impact on the outcome of public adherence to nonpharmaceutical measures and, ultimately, on the outcome of outbreak mitigation. Adding to the challenging of effective communications is the cross-propagation of relevant, and sometimes inconsistent, pieces of information that enter social media at different time points. This calls for a strategy of optimizing the timing of posting key information in social media during a fast-evolving pandemic.

Figure 1 shows the cross-propagation of three pieces of related information about COVID-19: titled as "Just want a regular 20200202," "Announcement of donation acceptance by Wuhan JinYinTan hospital," and "Three cases of community transmission were reported in Shenzhen for the first time." These pieces of information were posted in Sina Microblog, with different beginning and ending time points marked in the (horizontal) timeline. Almost immediately after reading information A, some users forwarded information B, so both pieces of information shared similar life cycle but with beginning and ending time points close to each other. There were 12,283 information A users, among which 742 (6.04%) forwarded information B. Eventually, 7161 of users forwarded information B, and those who simultaneously forwarded information A accounted for 10.36% (n=7161). Information C was then released, and users of information B started to forward information C. At the end, among 7161 users of information B, 1158 (16.17%) also forwarded information C, accounting for 10.26% (n=11,289) of the users for C.

Figure 1. Cross-propagation of three relevant pieces of information related to the COVID-19 pandemic, posted in sequence on Sina Microblog.



In general, relevant information, when posted with an appropriate time lag, can attract the interest of social media users in public hot events by increasing the efficiency of dissemination of a group of information cross-propagated. It is an important topic of research; the main objective of our study is to understand the information cross-propagation dynamics to inform optimal strategies of posting relevant information in an appropriate time sequence to ensure their maximum interaction for effective copromotion during a public health emergency situation.

To the best of our knowledge, no appropriate model framework has been developed and analyzed to examine the impact of information cross-propagation dynamics for a group of relevant information that is posted subsequently. Here, we try to fill this gap by proposing a susceptible-forwarding-immune model with

time-delayed posting and transmission. We developed and illustrated this framework, and parametrized our model by using the forwarding quantity that represents public attention to some popular opinions on the COVID-19 pandemic. Our focus is on the dynamic interactions among several pieces of information posted sequentially, and we aim to examine the impact of time lags between different posting time points on the evolution and steady states of cross-propagation.

In the field of information propagation dynamics, there is a strong similarity between the propagation of rumors and a pathogen's spread in the population [1]. Many studies have used epidemic models to examine rumor propagation in the hope that the negative influence of rumors can be eliminated or at least minimized. For example, the susceptible-infected-exposed-recovered (SIER) model [2-5], susceptible-infected model [6-9], susceptible-infected-susceptible [10,11] model, and susceptible-infected-recovered (SIR) model [12-14] have all been developed and recognized as classical propagation dynamics models.

The development of the internet and the enrichment of social media mandate further extensions of traditional models to reflect novel transmission mechanisms and to take advantage of data from multiple platforms. Gu and Cai [15] and Gu et al [16] proposed the forget-remember mechanism to study the spreading process in a 2-state model. Zhao et al [17] combined the forgetting mechanism and the SIR model to represent the rumor spreading process in an online social blogging platform LiveJournal. In 2014, Zhao et al [18] integrated the refutation mechanism in homogeneous social networks into the SIR model and analyzed the dynamic process of rumor propagation. Considering the three influencing factors of enterprises affected by rumors, pinion leaders, and a microblog platform, an SIR model based on browsing behavior was constructed to explain how rumors spread among followers under the influence of different rumor refuting measures [19]. Other features of the new media were further incorporated by Zhao et al [20]. Borge-Holthoefer et al [21] considered the case when spreaders were not always active and an ignorant was not interested in spreading the rumor, and then separately introduced these ideas into two different models. They concluded that these models provided higher adhesion to real data than classical rumor spreading models. In 2020, Yin et al [22] considered the user's behavior of re-entering new topics and proposed a multiple-information susceptible-discussing-immune model to investigate COVID-19–relevant information propagation in the Chinese Sina Microblog. Ding et al [23] proposed an improved SIR model, which used differential equations to study the rule of information transmission on media platforms and predicted microblog information accurately. Wang et al [24] proposed a modeling method that considers Weibo propagation behavior based on the susceptible-infected-susceptible model, so the forwarding trend in the future can be predicted. Zhang et al [25] focused on the impact of media transmission and interpersonal relationships on information propagation and then proposed the media and interpersonal relationship susceptible-infected-exposed-recovered model. Zhao et al [26] developed a new rumor spreading model called the susceptible-infected-hibernator-removed model, introducing a new kind of people-hibernators to reduce the maximum rumor influence. Woo et al [27] proposed an event-driven SIR model based on the impact of news releases on social media to reflect the impact of specific events on opinion diffusion. Yao et al [28] concentrated on examining the influence of different interactions among information on the spread of public opinion and modeling based on the SIR model, which verified the otherness of public opinion under a distinct information environment. For other studies relevant to our paper, see [29-32]. In particular, Tanaka et al [32] added a new module to the traditional model by using two data sets from the Japanese Mixi and Facebook rather than a single data set.

Many studies on cross-transmission in disease diffusion are highly significant to study the cross-propagation of information. Feng et al [33] established a mathematical model that incorporated the virus mutation dynamics in the transmission of the Chikungunya virus among mosquitoes and humans. However, the important phenomenon of time lag in posting and cross-propagation of relevant information for information dissemination in real social media networks has not been adequately addressed in these earlier studies. We noted that Zan [34] studied the double rumors spreading with different launch times, in which the new rumor was launched with a certain delay but also could interact with the old rumor. Zan [34] proposed two classes of double-rumors spreading models: a double-susceptible-infected-recovered model, where it was assumed that the rumor was disseminated by direct contacts of infective nodes with others, and a comprehensive double-susceptible-infected-recovered model, with which the authors studied the whole spread situation of all rumors with a focus on determining how many people did not spread all rumors in the entire period or how many were spreading or had spread at least one of the rumors.

In comparison with the aforementioned studies, here we consider the phenomenon where at different propagation stages of a piece of information posted in social media other pieces of information are posted, and their relevance in contexts and posting time sequence combined generate an outbreak for each piece of information and, more importantly, cross-propagation in which users of one piece of information forward other pieces of information they are exposed to later. We developed two classes of dynamic propagation models that focus on the single information transmission and multi-information cross-propagation patterns during explosive and quasi–steady-state periods of the information posted sequentially. We aim to examine emphatically the influence of different participating groups of posted information on the spread of the information from the participating groups. As populations who have forwarded, whether exposed or not exposed to relevant posted information, are attracted to a new piece of information differently, by introducing and analyzing the impact of *attractiveness indexes* on relevant information propagation and examining the significant factors of delaying in posting relevant information propagation, we inform strategies for sequentially posting relevant information to achieve effective communication of key public opinions. We will illustrate this

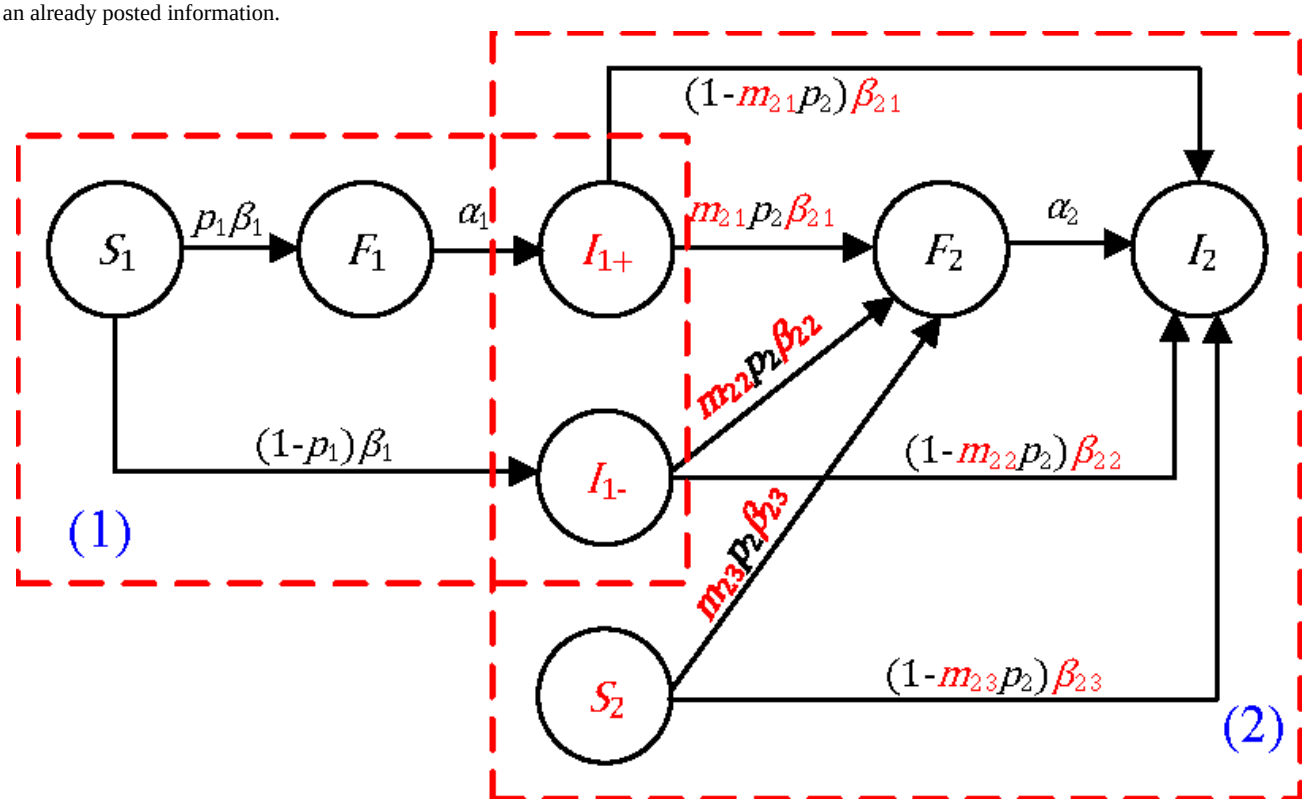
with data from public opinions about COVID-19 pandemic management.

Methods

Large Delay in Transmission Susceptible-Forwarding-Immune Dynamics Model

The structure of our proposed large interval delay in transmission susceptible-forwarding-immune (LTI DT-SFI) dynamics model is shown in Figure 2. There are two phases involved. In phase 1, a piece of stand-alone information (information 1) is spreading; and in phase 2, another piece of information (information 2) is posted during the quasi–steady-state period of the posted information.

Figure 2. A schematic illustration of the information cross-spreading, where a new piece of information is posted during the quasi steady-state period of an already posted information.



Phase 1: Post a Piece of Stand-alone Information

The propagation dynamics during phase 1 for one piece of posted information is modeled based on the traditional susceptible-forwarding-immune [35] model, with a novel stratification of the immune population. Namely, there will be two classes of immune populations (as far as information 1 is concerned): those who have forwarded the posted information but are no longer in the active period of forwarding this posted information (I_{1+}) and those who have been exposed to the information but are not interested in forwarding it (I_{1-}). This distinction of immunity is important, as individuals in these two distinct compartments will have different levels of interest in other relevant information that will be posted later. This will allow us to introduce different measures of attractiveness to new relevant information.

So, in our model, we stratify the population (N_1) into four states: the susceptible state of the posted information (S_1), the forwarding state of the posted information (F_1), the inactive immune state (I_{1+}), and the direct immune state (I_{1-}). A susceptible user can be exposed to the posted information with an average exposure rate β_1 and will forward the information with the forwarding probability p_1. The forwarding users can become inactive immune users with an average rate α_1. So a user may have a unique state, with $S_1(t)$, $F_1(t)$, $I_{1+}(t)$, and $I_{1-}(t)$ denoting the number of users in the susceptible, forwarding, inactive, and direct immune state, respectively. We obtain the following delay in transmission susceptible-forwarding-immune (DT-SFI) dynamics model in phase 1:

The state transition of different populations in phase 1 can be interpreted as follows: an active forwarding user will contact an average number of $\beta_1 N_1$ users per unit time, and the probability of a user being a susceptible user of the posted information is $S_1(t)/N_1$, so an active forwarding user will contact $\beta_1 S_1(t)$ susceptible users. There are $F_1(t)$ active forwarding users of information 1 in total at time t, so $p_1\beta_1 S_1(t)F_1(t)$ susceptible users will choose to forward the information and become new forwarding users, and $(1 - p_1)\ \beta_1 S_1(t)F_1(t)$ will not. As time goes by, $\alpha_1 F_1(t)$ will go to the immune state from the forwarding

period when they do not influence other users as far as information 1 is concerned.

For data fitting purposes, we noted that the Sina Microblog provides important data about any piece of information relevant to COVID-19, the number of cumulative forwarding quantity, given by:

Phase 2: Post New Information During the Quasi–Steady-state Period of the Posted Information

Now we consider that a piece of new information is posted at time t_τ when the posted information is already in the quasi–steady state. We introduce the following indexes:

- *An extensive exposure attractiveness index*: for the individuals in the immune state who have forwarded the posted information but are no longer in their active forwarding period. These individuals are more susceptible to the new yet relevant information, as they had interest in the posted information.
- *A mild exposure attractiveness index*: for the direct immune individuals. These individuals have had exposure to the posted information but had shown little interest in the information.
- *An unexposure attractiveness index*: for those who were never exposed to the first posted information

Accordingly, we introduce three states of the population (N_2) for the newly posted information: the susceptible state (S_2), the forwarding state (F_2), and the immune state (I_2). We summarize the notations in Textbox 1.

Textbox 1. Parameter definitions for large delay model.

Attractiveness parameters, stratified by the exposure to old information

- m_{21}: The extensive exposure attractiveness index that an inactive user of state I_{1+} becomes a forwarding user of state F_2.
- m_{22}: The mild exposure attractiveness index that a direct immune user of state I_{1-} becomes a forwarding user of state F_2.
- m_{23}: The unexposure attractiveness index that a new susceptible user of state S_2 becomes a forwarding user of state F_2.

Transmission parameters associated with the different attractiveness

- β_{21}: The average exposure rate that the inactive users of the old information can contact with the newly posted information.
- β_{22}: The average exposure rate that the direct immune users of the old information can contact with the newly posted information.
- β_{23}: The average exposure rate that the new susceptible users can contact with the newly posted information.
- p_2: The probability that an exposed user will forward the newly posted information.
- α_2: The average rate that a user in the forwarding state of newly posted information becomes inactive to forwarding, where $1/\alpha_2$ is the average duration a forwarding user remains active in forwarding newly posted information.

Each user may have a unique state, with $I_{1+}(t)$, $I_{1-}(t)$, $S_2(t)$, $F_2(t)$, and $I_2(t)$ denoting the number of users in the susceptible, forwarding, and immune state, respectively. We obtain the following LTI DT-SFI dynamics model in phase 2:

The mass action in phase 2 can be interpreted as follows: an active forwarding user will contact an average number of $\beta_{21}N_2$ inactive immune users of the posted information per unit time, and the probability of a user being an inactive immune user is $I_{1+}(t)/N_2$, so an active forwarding user will contact $\beta_{21}I_{1+}(t)$ inactive immune users, among which $m_{21}p_2\beta_{21}I_{1+}(t)F_2(t)$ will choose to forward the new information and $(1 - m_{21}p_2)\beta_{21}I_{1+}(t)F_2(t)$ will not, where $F_2(t)$ is the number of new active forwarding users at time t; an active forwarding user will contact an average number of $\beta_{22}N_2$ direct immune users of the posted information per unit time, and the probability of a user being a direct immune user is $I_{1-}(t)/N_2$, so an active forwarding user will contact $\beta_{22}I_{1-}(t)$ direct immune users, among which $m_{22}p_2\beta_{22}I_{1-}(t)F_2(t)$ will choose to forward the new information and $(1 - m_{22}p_2)\beta_{22}I_{1-}(t)F_2(t)$ will not, where $F_2(t)$ is the number of new active forwarding users at time t; an active forwarding user will contact an average number of $\beta_{23}N_2$ susceptible users of the newly posted information per unit time, and the probability of a user being a susceptible user is $S_2(t)/N_2$, so an active forwarding user will contact $\beta_{23}S_2(t)$ susceptible users, among which $m_{23}p_2\beta_{23}S_2(t)F_2(t)$ will choose to forward the new information and $(1 - m_{23}p_2)\beta_{23}S_2(t)F_2(t)$ will not, where $F_2(t)$ is the number of new active forwarding users at time t.

The forwarding quantity of the newly posted information is given by:

The Public Opinion Reproduction Ratio

Since the newly posted Weibo starts at different times and develops differently under the influence of prior information, we defined the information reproduction ratio as:

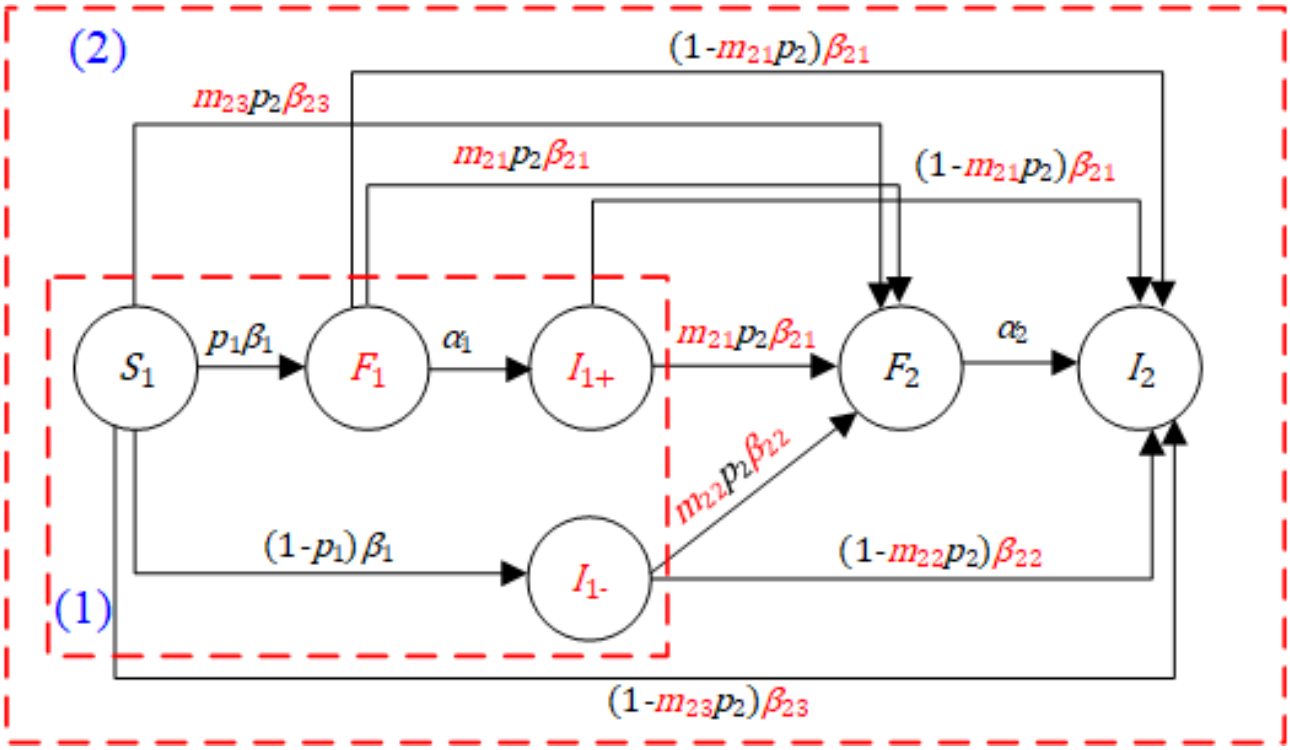
This is the total number of information 2 users generated by introducing a typical information 2 user, at time τ after

information 1 was posted, during their entire period of active forwarding. The initial population for information 2 has been stratified by the exposure of the entire population to information 1 during the time interval [0, τ]. The relative size of [x] to the unity determines if information 2 can generate an information outbreak; since [x] (0) = ($m_{21}p_2\beta_{21}I_{1+,\tau} + m_{22}p_2\beta_{22}I_{1-,\tau} + m_{23}p_2\beta_{23}S_{20} - \alpha_2$) F_2 (0), we concluded that [x] (0)>0 if [x]>1.

Short Delay in Transmission Susceptible-Forwarding-Immune Dynamics Model

Our comprehensive short interval delay in transmission susceptible-forwarding-immune (STI DT-SFI) dynamics model based on the forwarding quantity is shown in Figure 3. In this model, we include two phases. In phase 1, a stand-alone piece of information (information 1) is spreading, corresponding to phase 1 in the LTI DT-SFI model. In phase 2, a piece of newly posted information (information 2) is posted at t_τ during an outbreak period of the posted information. Here, we also divide the population into three groups: S population (S), F population (F_1, F_2), and I population (I_{1+}, I_{1-}, I_2). In particular, we think of the susceptible state of both the posted information and the new information as a whole (S). t_τ is the post time of the newly posted information.

Figure 3. A schematic diagram to illustrate information spreading, when the post time of the newly posted information is during the outbreak period of the posted information.



Phase 1: Post Stand-alone Information

The model is consistent with that of phase 1 of the LTI DT-SFI model.

Phase 2: Posting New Information During the Outbreak Period of the Posted Information

Considering that the new information is posted during the outbreak period of the old information, we developed our phase 2 model to describe the concurrent dynamic process of the two related pieces of information. We considered the difference between the population in active forwarding state or the immune state out of the active period and the population in the direct immune state of the first (old) piece of information. Here, we set the following indexes:

- *An extensive exposure attractiveness index*: for the individuals in the forwarding state, who have forwarded the posted information but are still in their active forwarding period, and the individuals in the immune state, who have forwarded the old information but are no longer in their active forwarding period, to indicate that this population will be attracted by the new information due to the relevance of the two pieces of information
- *A mild exposure attractiveness index*: for the direct immune population to portray that the population will be attracted due to a moderate contact
- *An unexposure attractiveness index*: for the integrated susceptible population to describe that the population will be attracted when they have never read the related information

Assuming that the number of users (N_3) who can contact the information in the process of information propagation on Sina Microblog remains unchanged, we introduced three states of the population of the newly posted information: the susceptible state (S_1) that includes the users who can be exposed to the old information and the new information, the forwarding state of

the new information (F_2), and the immune state (I_2). The parameters are shown in Textbox 2.

Textbox 2. Parameters definitions for short delay model.

Attractiveness parameters, stratified by the exposure to old information

- m_{21}: The extensive exposure attractiveness index that an inactive user of state I_{1+} becomes a forwarding user of state F_2.
- m_{22}: The mild exposure attractiveness index that a direct immune user of state I_{1-} becomes a forwarding user of state F_2.
- m_{23}: The unexposure attractiveness index that a new susceptible user of state S_2 becomes a forwarding user of state F_2.

Transmission parameters associated with the different attractiveness

- β_1: The average exposure rate that the susceptible users can contact the first information.
- β_{21}: The average exposure rate that the inactive users of the old information can contact the newly posted information.
- β_{22}: The average exposure rate that the direct immune users of the old information can contact the newly posted information.
- β_{23}: The average exposure rate that the new susceptible users can contact the newly posted information.
- p_1: The probability that the susceptible users will forward the first information.
- p_2: The probability that the exposed users will forward the newly posted information.
- α_2: The average rate that a user in the forwarding state of newly posted information becomes inactive to forwarding, where $1/\alpha_2$ is the average duration a forwarding user remains active in forwarding newly posted information.

Each user may have a unique state, with $S_1(t)$, $F_1(t)$, $I_{1+}(t)$, $I_{1-}(t)$, $F_2(t)$, and $I_2(t)$ denoting the number of users in the susceptible, forwarding, and immune state when $t>0$, respectively. We obtain the following STI DT-SFI dynamics model in phase 2:

The mass action in phase 2 can be interpreted as follows: an active forwarding user will contact an average number of $\beta_{21}N_3$ inactive immune users and forwarding users of posted information per unit time, and the probability of a user being an inactive immune user and a forwarding user are $I_{1+}(t)/N_3$ and $F_1(t)/N_3$, respectively, so an active forwarding user will contact $\beta_{21}I_{1+}(t)$ inactive immune users and $\beta_{21}F_1(t)$ forwarding users, among which $m_{21}p_2\beta_{21}I_{1+}(t)F_2(t)$ and $m_{21}p_2\beta_{21}F_1(t)F_2(t)$ will choose to forward the new information, however, $(1 - m_{21}p_2)\beta_2 I_{1+}(t)F_2(t)$ and $(1 - m_{21}p_2)\beta_{21}I_1(t)F_2(t)$ will not, where $F_2(t)$ is the number of new active forwarding users at time t. Here, the individuals in the forwarding state and the individuals in the immune state have the same familiarity with the topic content who have forwarded information 1; an active forwarding user will contact an average number of $\beta_{22}N_3$ direct immune users of posted information per unit time, and the probability of a user being a direct immune user is $I_{1-}(t)/N_3$, so an active forwarding user will contact $\beta_{22}I_{1-}(t)$ direct immune users, among which $m_{22}p_2\beta_{22}I_{1-}(t)F_2(t)$ will choose to forward the new information and $(1 - m_{22}p_2)\beta_{22}I_{1-}(t)F_2(t)$ will not, where $F_2(t)$ is the number of new active forwarding users at time t; an active forwarding user will contact an average number of $\beta_{23}N_3$ susceptible users, and the probability of a user being a susceptible user is $S_1(t)/N_3$, so an active forwarding user will contact $\beta_{23}S_1(t)$ susceptible users, among which $m_{23}p_2\beta_{23}S_1(t)F_2(t)$ will choose to forward the new information and $(1 - m_{23}p_2)\beta_{23}S_1(t)F_2(t)$ will not, where $F_2(t)$ is the number of new active forwarding users at time t.

The forwarding quantity of the newly posted information is given by:

The Public Opinion Reproduction Ratio

Considering the initial condition in phase 2, we can obtain the following public opinion reproduction ratio  . The new information entered at time τ during an outbreak period of the posted information, and  $(0) = (m_{21}p_2\beta_{21}F_{1\tau} + m_{21}p_2\beta_{21}I_{1+,\tau} + m_{22}p_2\beta_{22}I_{1-,\tau} + m_{23}p_2\beta_{23}S_{1\tau} - \alpha_2)F_2(0)$. The population will never take off if  $(0) = (m_{21}p_2\beta_{21}F_{1\tau} + m_{21}p_2\beta_{21}I_{1+,\tau} + m_{22}p_2\beta_{22}I_{1-,\tau} + m_{23}p_2\beta_{23}S_{1\tau} - \alpha_2)F_2(0)<0$ due to the decrease of $S_{10\tau}$. It is therefore natural to introduce the following as the STI DT-SFI reproduction ratio:

In the same way, the  of the STI DT-SFI model denotes the comprehensive public opinion generated by the newly posted Weibo starting at the outbreak period of the posted information. When the reproduction ratio <1, it means that the new public opinion will decline. When the reproduction ratio >1, it indicates that the new public opinion will initially grow exponentially.

Statistical Analysis

Data Description

Since the COVID-19 outbreak in China, intensive information that were clearly relevant to each other have been frequently posted. Figure 4 shows the total forwarding quantities of Weibos

during each 1-hour time frame on January 25, 2020, and January 26, 2020 (data acquisition up to February 19, 2020), for the top 10 opinion leaders of this outbreak event in the Chinese Sina Microblog. As illustrated, those pieces of information with high influence were posted frequently by each opinion leader. At the same time, there was a strong correlation among a series of Weibos posted by these opinion leaders during certain periods. Of all the data shown in Figure 4, People's Daily issued five Weibos in 1 hour from 10 PM to 11 PM on January 25, and they were forwarded by a total of more than 300,000 users. Therefore, the frequent release of relevant information by original post owners was a common phenomenon in the COVID-19 information propagation, and understanding its effectiveness is important.

Figure 5 shows the trend of cumulative forwarding users for the three pieces of information in Tables 1-3. It shows that when information A broke out, information B was posted almost immediately. Compared with the information, the outbreak period of information B was shorter and the trend was flatter. Information C was released during the quasi–steady-state period of information B. In comparison with this information, the outbreak period of information C lasted longer; meanwhile, the cumulative forwarding quantity was also larger.

Sequentially releasing two related pieces of information by the same original post owners within the same COVID-19 theme was a common phenomenon. Importantly, different entering times of new information during the spreading process of an old (previously posted) information exhibited different promoting effects on the cross-propagation and cross-promotion of relevant public opinions. Here we focus on users who have been exposed to one posted information that may have a special interest in, and hence are susceptible to, new and relevant information. This represents a remarkable difference from the spread of rumors and other traditional public hot events. Our information cross-propagation DT-SFI models, including the STI DT-SFI dynamics model and the LTI DT-SFI dynamics model, were developed to take into consideration the situations when the relevant information is posted during the outbreak period or during the quasi–steady-state period of the previously posted (old) information.

Figure 4. A bubble chart of forwarding quantity of public opinions on COVID-19 information by top ten opinion leaders in Weibos, during an early period of COVID-19 outbreak in China.

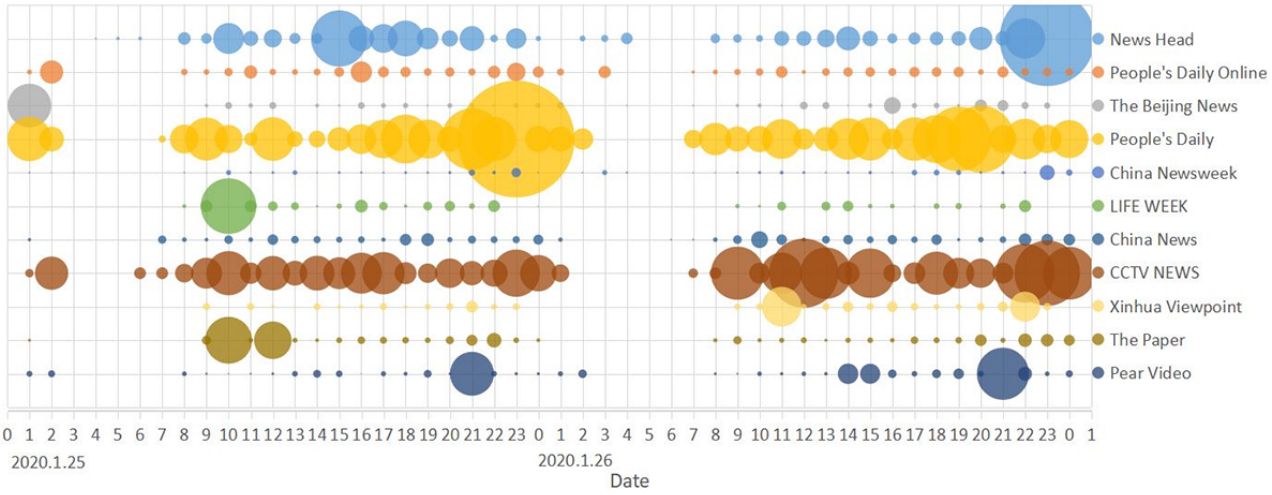


Figure 5. Cumulative forwarding quantity of three pieces of information.

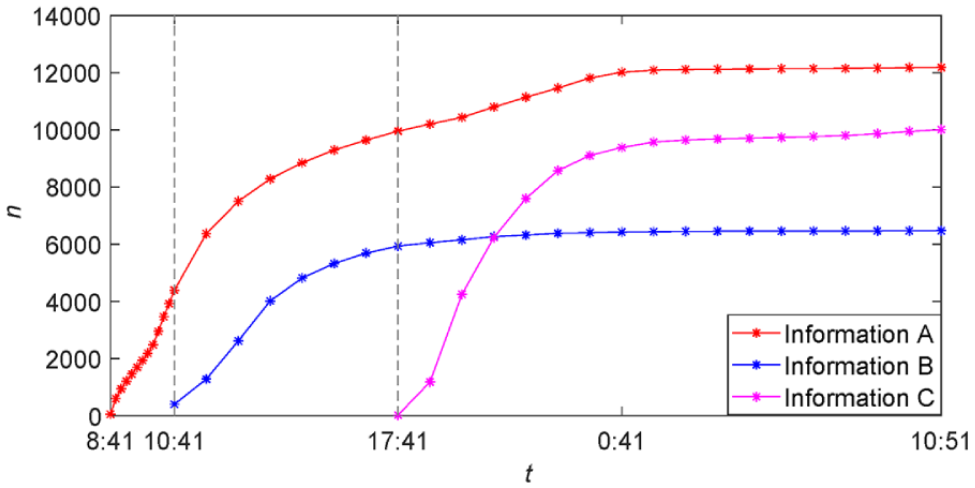


Table 1. Cumulative forwarding quantity of information A posted at 8:41 AM on February 2, 2020.

Time	Information A forwarding quantity, n
0 min	47
10 min	597
20 min	940
30 min	1208
40 min	1458
50 min	1691
60 min	1937
70 min	2182
80 min	2477
90 min	2952
100 min	3461
110 min	3917
2 h	4390
3 h	6366
4 h	7501
5 h	8281
6 h	8846
7 h	9293
8 h	9638
9 h	9954
10 h	10,199
11 h	10,435
12 h	10,795
13 h	11,138
14 h	11,459
15 h	11,812
16 h	12,013
17 h	12,088
18 h	12,109
19 h	12,119
20 h	12,128
21 h	12,136
22 h	12,140
23 h	12,146
24 h	12,157
25 h	12,171
26 h	12,184

Table 2. Cumulative forwarding quantity of information B posted at 10:41 AM on February 2, 2020.

Time (hours)	Information B forwarding quantity, n
2	15
3	1281
4	2615
5	4013
6	4817
7	5322
8	5685
9	5932
10	6052
11	6152
12	6264
13	6317
14	6380
15	6401
16	6423
17	6434
18	6447
19	6454
20	6455
21	6456
22	6458
23	6460
24	6461
25	6465
26	6471

Table 3. Cumulative forwarding quantity of information C posted at 5:51 PM on February 2, 2020.

Time (hours)	Information C forwarding quantity, n
9	20
10	1180
11	4244
12	6235
13	7595
14	8572
15	9103
16	9381
17	9569
18	9642
19	9680
20	9700
21	9736
22	9764
23	9800
24	9864
25	9943
26	10,006

Data Fitting for the LTI DT-SFI Model

Parameter Estimation

To fit our model with real data from the Sina Microblog, we used the least squares (LS) method to estimate the LTI DT-SFI model parameters and the initial data. In phase 1, the parameter vector is set as $\Theta_1 = (p_1, \beta_1, \alpha_1, S_{10})$, and the corresponding numerical calculation based on the parameter vector for $C_1(t)$ is denoted by (k, Θ_1). The follow LS error function was used in our calculation:

where C_{1k} denotes the actual cumulative forwarding populations of the posted information. Similarly, in phase 2, the vector is set as $\Theta_2 = (\beta_{21}, \beta_{22}, \beta_{23}, m_{21}, m_{22}, m_{23}, p_2, \alpha_2, S_{20})$, and the corresponding numerical calculation based on the parameter vector for $C_2(t)$ is denoted by (k, Θ_2). The following LS error function was used in our calculation:

where C_{2k} denotes the actual cumulative forwarding populations of the newly posted information. Here, n=1, 2 ... represents the different phases, and k=0, 1, 2, ... is the sampling time n=1, 2, 3. We estimated the parameters of our *LTI DT-SFI* model with the data of information B and information C.

Figure 6 reports our data fitting results for information B and information C on the real data given in Tables 2 and 3, where the blue star denotes the actual cumulative number of forwarding users of information B; the red star denotes the actual cumulative number of forwarding users of information C; and the green line and the black line denote the estimated cumulative number of forwarding users of information B and information C, respectively.

Tables 4 and 5 give estimated values of important parameters for information B and information C, respectively. We can see in phase 2, when information C was posted during the quasi–steady-state period of information B, the average exposure rate β_{21} was the largest, indicating that an inactive user of the posted information B was more susceptible to the newly posted information C; the average exposure rate β_{23} was small, indicating that a susceptible user of the newly posted information C contacted the information at a lower rate. In addition, among the three attractiveness indexes, the index m_{22} is the largest, which indicates that information C had the strongest appeal to a direct immune user of information B and has the least attractiveness to an inactive user of the posted information B.

By comparison, there is a difference between the initial time of a new piece of information at the outbreak phase and at the quasi–steady-state phase of the posted information. When the initial time of new information is in the outbreak phase of the posted information, the value of the average immune rate α_2 is generally higher than the value in the quasi–steady-state phase, which is due to the rapid outbreak of information (a large amount of information updates and iterations). The average active duration $1/\alpha_2$ of forwarding users of the new piece of information where users can influence other users to contact information is shorter. Similarly, the average forwarding

probability p_2 of the initial time in the outbreak period is also higher than the value in the quasi–steady-state phase, which conforms to the fact that people are more willing to participate in the discussion successively when exposed to relevant information in the short term. In comparison, the value of average contact rate β_{21} in the outbreak phase is lower than that in the quasi–steady-state phase, indicating that the population who has forwarded information will be larger based on a relatively larger contact population. β_{22} and β_{23} of the initial time in the outbreak period are higher than those of the initial time in the quasi–steady-state phase, indicating that continuous exposure to relevant information in the short term (when the initial time of the new piece of information is in the outbreak phase of a posted information) would attract people who had not participated in the new transmission to forward and spread the information. All average attractiveness indexes of the initial time in the quasi–steady-state phase are larger, which indicates that information that is re-exposed to users after a period of time will inspire their freshness and make them pay more attention to the information itself.

Figure 6. The data fitting results of Information B and Information C.

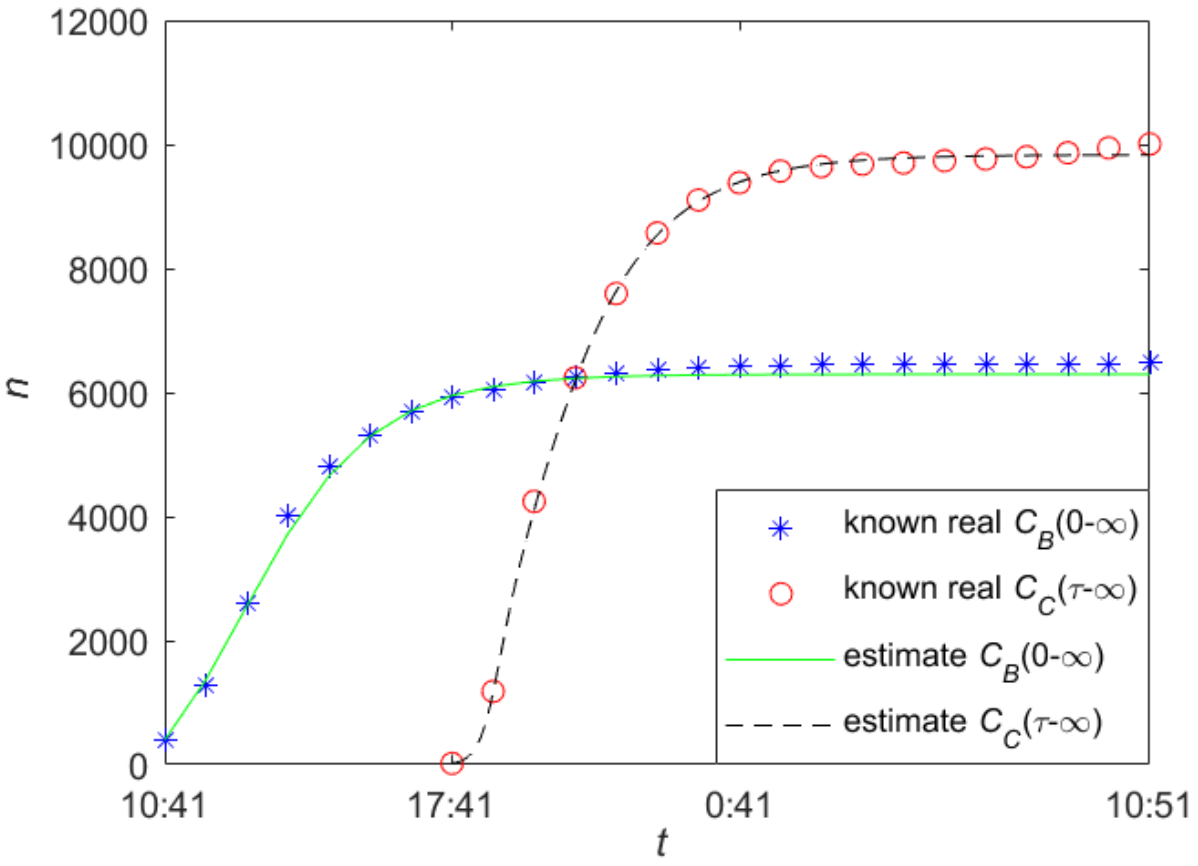


Table 4. Values of some important parameters, estimated for information B.

Parameter	Estimated value	Standard error	Minimum	Maximum
S_{10}	5.6458×10^{6}	81.0304	0.0000	1.0000×10^{8}
α_1	1.5757	0.0427	0.0000	4.0000
β_1	1.7901×10^{-4}	1.7463×10^{-5}	0.0000	1.0000
p_1	0.0020	1.5023×10^{-4}	0.0000	1.0000

Table 5. Values of important parameters, estimated for information C.

Parameter	Estimated value	Standard error	Minimum	Maximum
S_{20}	7.4439×10^{6}	1.3770×10^{3}	0.0000	1.0000×10^{8}
α_2	0.9858	0.0993	0.0000	4.0000
β_{21}	0.8994	0.2829	0.0000	1.0000
β_{22}	0.0023	0.0023	0.0000	1.0000
β_{23}	1.0871×10^{-4}	1.2088×10^{-4}	0.0000	1.0000
m_{21}	0.2468	0.1132	0.0000	2.0000
m_{22}	1.9895	0.1087	0.0000	2.0000
m_{23}	0.5559	0.2646	0.0000	2.0000
p_2	2.9516×10^{-4}	1.8597×10^{-4}	0.0000	1.0000

Data Fitting for the STI DT-SFI Model

Parameter Estimation

To use our model to explore some distinctions of the qualitative behaviors for prediction, we used the LS method to estimate the STI DT-SFI model parameters and the initial data of our model. The vector is set as $\Theta_3 = (p_1, \beta_1, \alpha_1, p_2, \beta_{21}, \beta_{22}, \beta_{21}, m_{21}, m_{22}, m_{23}, \alpha_2, S_{10})$, and the corresponding numerical calculation based on the parameter vectors for $C_1(t)$ and $C_2(t)$ are denoted by (k, Θ_3) and (k, Θ_3), respectively. The following LS error function was used in our calculation:

where C_{1k} and C_{2k} denote the actual cumulative forwarding populations of the posted information and the newly posted information; here, n=1, 2, 3 represents the different phases, and k=0, 1, 2, ... is the sampling time n=1, 2, 3. We estimated the parameters of our *STI DT-SFI* model with the data of information A and information B.

In the data fitting of the *STI DT-SFI* model, we used the same method as for the LTI DT-SFI model to fit the data of information A and information B. As shown in Figure 7, we performed data fitting of information A and information B on the real data in Tables 1 and 2, where the pink star denotes the actual cumulative number of forwarding users of information A; the red star denotes the actual cumulative number of forwarding users of information B; the green line and the blue line denote the estimated cumulative number of forwarding users in the early and later period of information A, respectively; and the black line denotes the estimated cumulative number of forwarding users of information B. It can be seen that our STI DT-SFI model achieves accurate estimation.

Table 6 gives some important values of parameter (relevant to the early period of the outbreak) estimation for information A, and Table 7 gives some important values of parameter estimation for the later period data of information A and all data of information B. We can see in phase 2, when information B was posted during the outbreak period of information A, the average exposure rate β_{21} and β_{22} are much larger than β_1 and β_{23}, which indicates that users who have been exposed to information A will contact information B at a greater rate than new susceptible users. In addition, the unexposure attractiveness index m_{23} is the largest among the three attractiveness indexes since the time interval between two information posts is small, and people who have not been exposed to relevant information may have a greater interest in new information; the outbreak of information B has the strongest appeal to susceptible users.

Figure 7. The data fitting results of Information A and Information B.

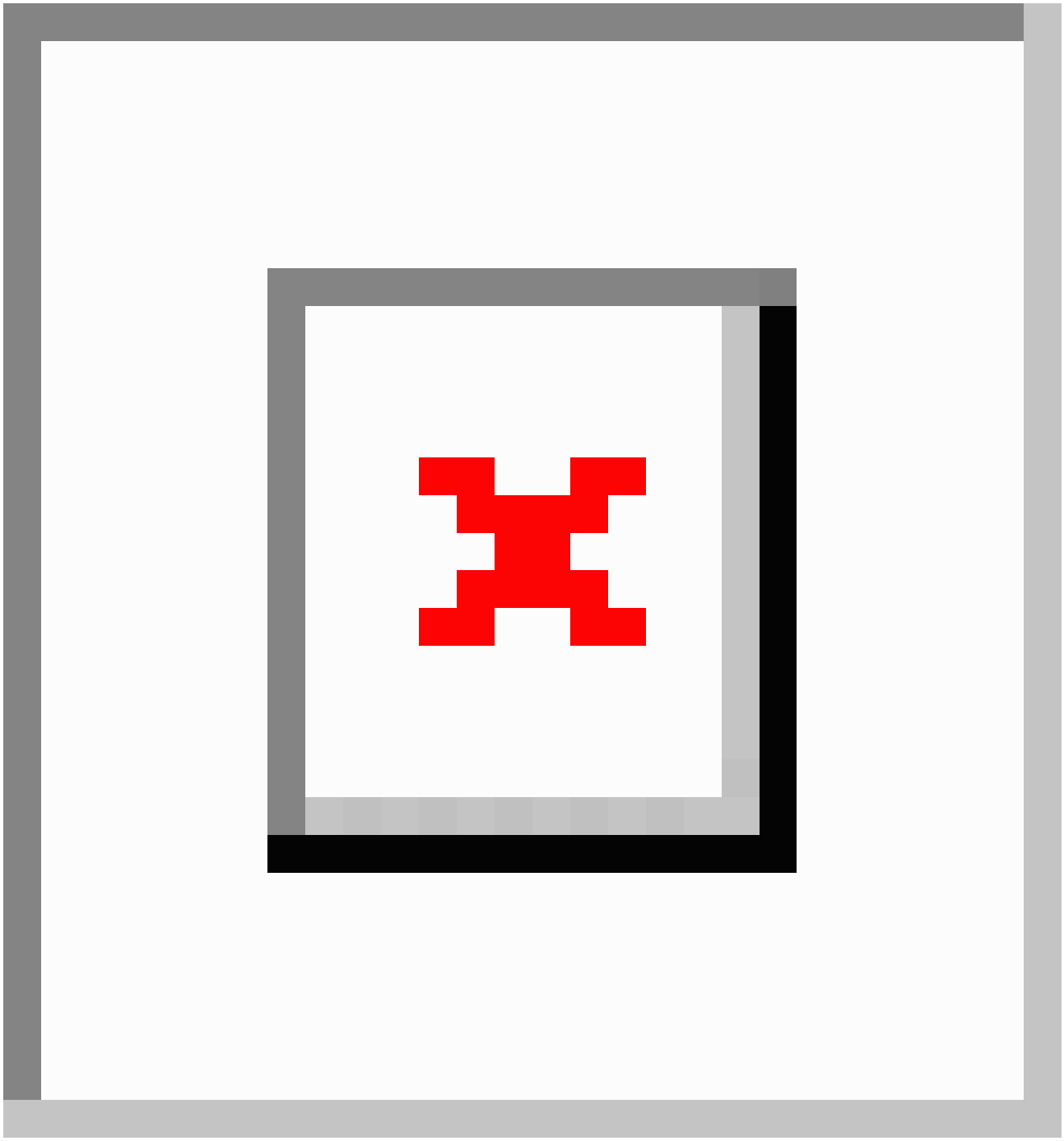


Table 6. Values of some important parameters, estimated for information A.

Parameters	Estimated value	Standard error	Minimum	Maximum
S_{10}	5.1682×10^{4}	28.3841	0.0000	1.0000×10^{7}
α_1	3.9986	0.4214	0.0000	4.0000
β_1	8.2700×10^{-5}	1.5673×10^{-5}	0.0000	1.0000
p_1	0.9823	0.1543	0.0000	1.0000

Table 7. Values of important parameters, estimated for information B.

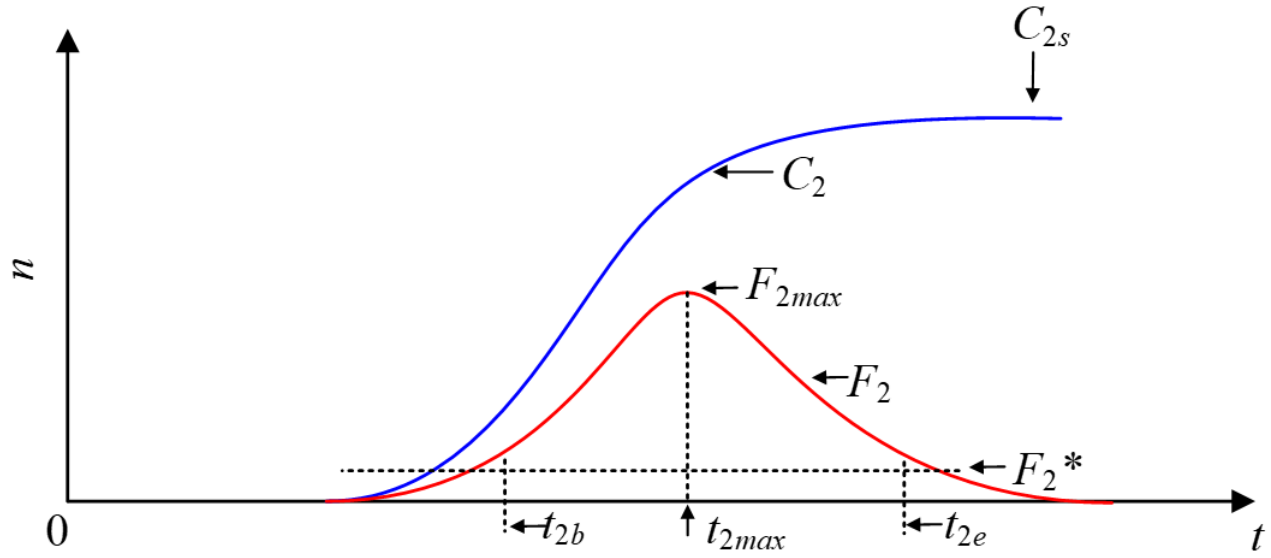
Parameter	Estimated value	Standard error	Minimum	Maximum
S_{10}	2.1494×10^{6}	208.4607	1.0000×10^{5}	1.0000×10^{7}
α_1	3.4777	0.1360	2.5000	3.50000
α_1	1.9159	0.0706	1.5000	3.50000
β_1	3.6601×10^{-4}	1.1663×10^{-4}	0.0000	4.0000×10^{-4}
β_{21}	0.0037	7.4426×10^{-4}	0.0000	0.0040
β_{22}	0.8184	0.0932	0.0000	1.0000
β_{23}	6.8834×10^{-5}	3.6532×10^{-5}	0.0000	$1.00007.4426 \times 10^{-4}$
m_{21}	0.0406	0.0196	0.0000	0.2000
m_{22}	0.0109	0.0093	0.0000	0.2000
m_{23}	0.1868	0.0445	0.0000	0.2000
p_1	0.0091	0.0026	0.0000	0.0200
p_2	0.0788	0.0346	0.0000	0.2000

Results

Influencing Factors Analytics: Information Release and Dissemination for the LTI DT-SFI Model

To make a qualitative and quantitative analysis of the delay in transmission, we introduced some additional indexes, shown in Figure 8, and show how these can be used to characterize the cross-propagation. We considered different effects of the posted information on newly posted information when the posted information has reached a quasi–steady state or is still in its outbreak period.

Figure 8. Some summative indices of a newly posted information that cross-propagating with an old information.



- *The outbreak peak F_{2max}*: the maximum of curve F_2, which reflects the peak user values of the newly posted information
- *The final size C_{2S}*: the stable state of curve C_2, which gives the final size of the total number of users of the newly posted information
- *The outbreak time t_{2b}, the end time t_{2e}, and the duration t_{2i}*: the definition depends on the outbreak threshold F_2* set in advance so that $F_2(t_{2b}) = F_2{*}=F_2(t_{2e})$. Here, t_{2b} denotes the outbreak time of the newly posted information, t_{2e} denotes the end time, and $t_{2i} = t_{2e}–t_{2b}$ denotes the duration of the newly posted information transmission. These time indexes will help us judge the start and end of the newly posted information transmission.
- *The outbreak velocity V_{2o} and the declining velocity V_{2d}*: the definition depends on $V_{2o}= (F_{2max}–F_2{*}) / (t_{2max}–t_{2b})$ and $V_d= (F_{2max}–F_2{*}) / (t_{2e}–t_{2max})$ when $F_2(t)=F_{2max}$ and t_{2max} is definite, which reflects the speed of the outbreak and the decline of the newly posted information.

To further analyze the different parameters responsible for the LTI DT-SFI model, we performed an analysis of partial rank

correlation coefficients [36] to evaluate the sensitivity based on 1000 samples for various input parameters against the threshold condition. According to the histogram and scatter diagram of dependence, when the correlation is positive, it means that, with the increase of the value of the parameter, the corresponding index value will increase; on the contrary, when the correlation is negative, the index will decrease as the parameter decreases. Figures 9-12 give the partial rank correlation coefficient results and partial rank correlation coefficient scatterplots with indexes , F_{2max}, $C_{2\infty}$, t_{2b}, t_{2i}, t_{2max}, V_{2o}, and V_{2d} with nine parameters (β_{21}, β_{22}, β_{23}, p_2, α_2, m_{21}, m_{22}, m_{23}, and S_{20}) of the newly posted information in the LTI DT-SFI model, respectively.

Figure 9. PRCC results and PRCC scatterplots with indexes for different parameters of the newly posted information in the large interval delay in transmission susceptible-forwarding-immune model. PRCC: partial rank correlation coefficient.

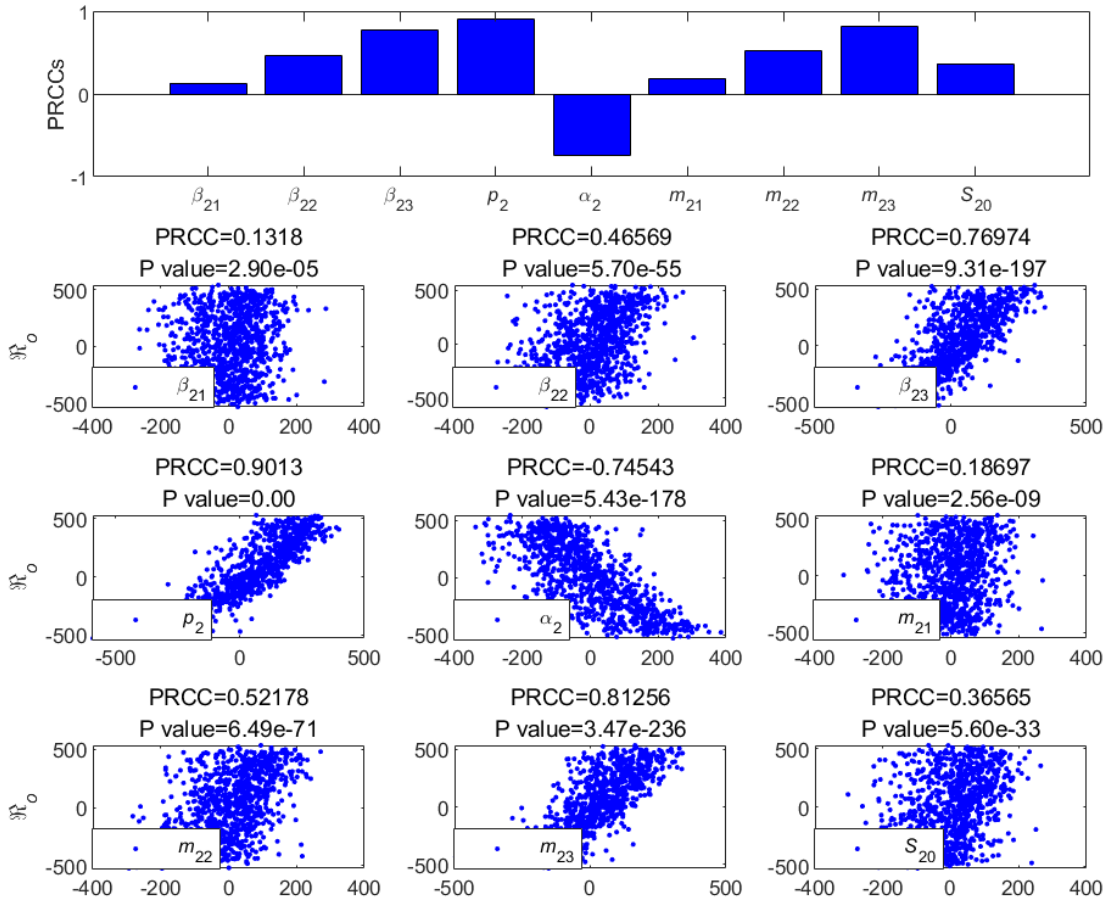


Figure 10. PRCC results and PRCC scatterplots with indexes F_{2max} and $C_{2\infty}$ for different parameters of the newly posted information in the large interval delay in transmission susceptible-forwarding-immune model. PRCC: partial rank correlation coefficient.

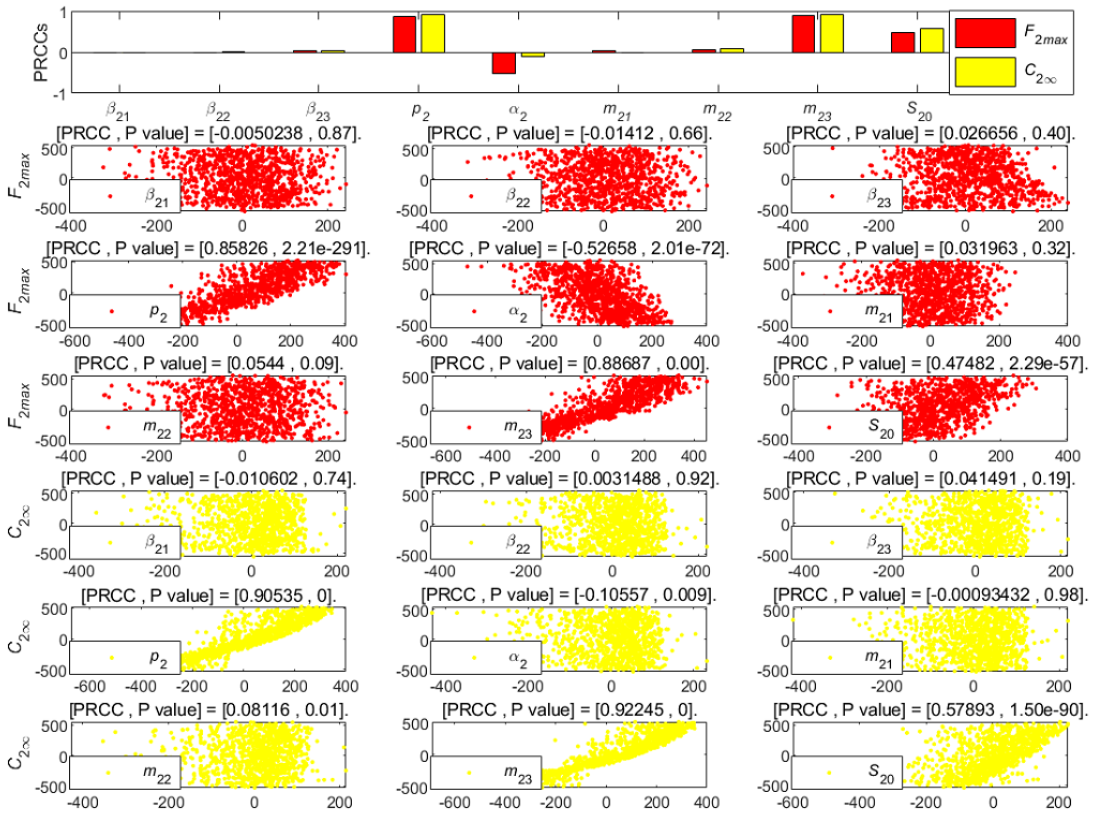


Figure 11. PRCC results and PRCC scatterplots with indexes t_{2max}, t_{2b}, and t_{2i} for different parameters of the newly posted information in the large interval delay in transmission susceptible-forwarding-immune model. PRCC: partial rank correlation coefficient.

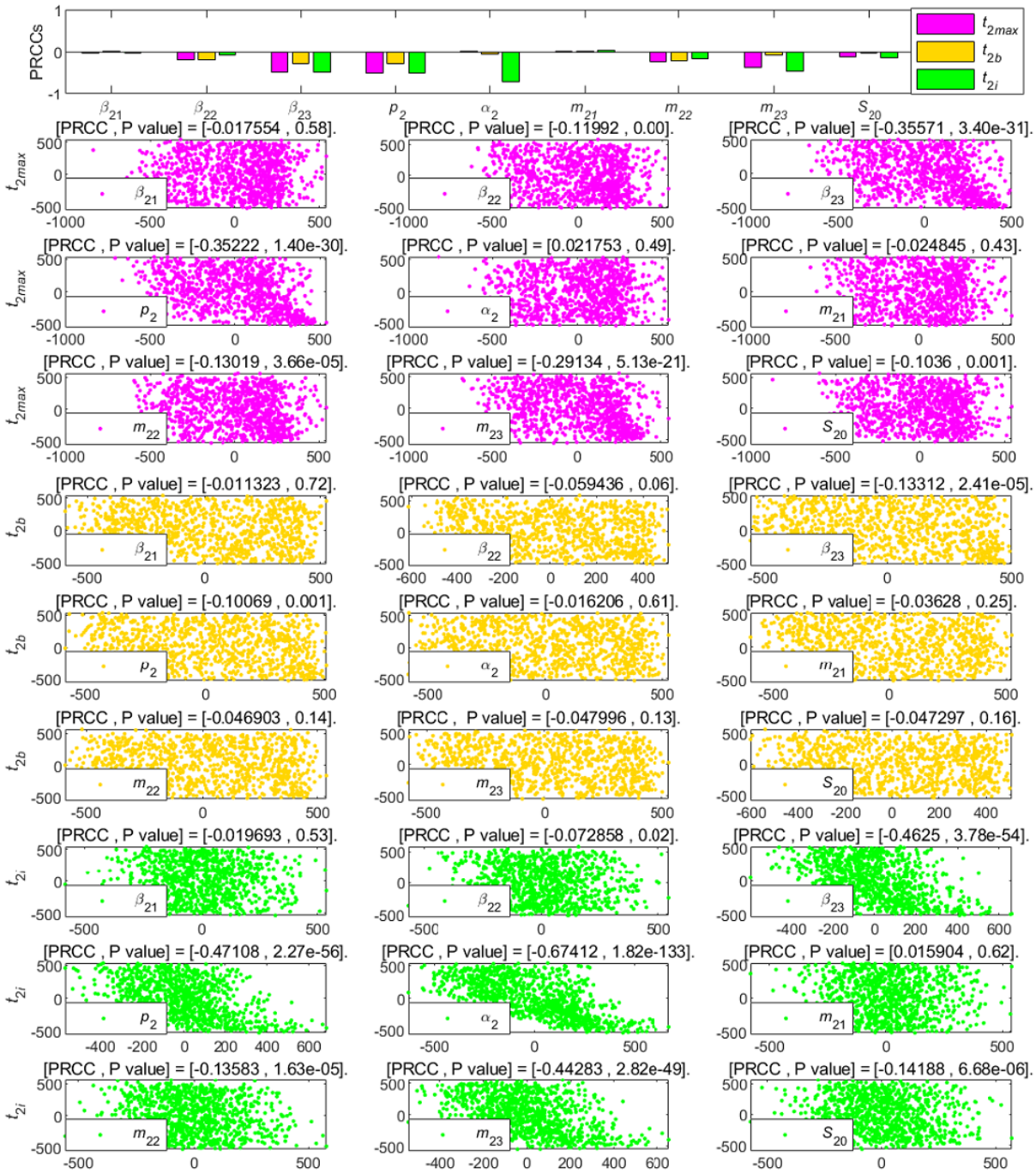


Figure 12. PRCC results and PRCC scatterplots with indexes V_{2o} and V_{2d} for different parameters of the newly posted information in the large interval delay in transmission susceptible-forwarding-immune model. PRCC: partial rank correlation coefficient.

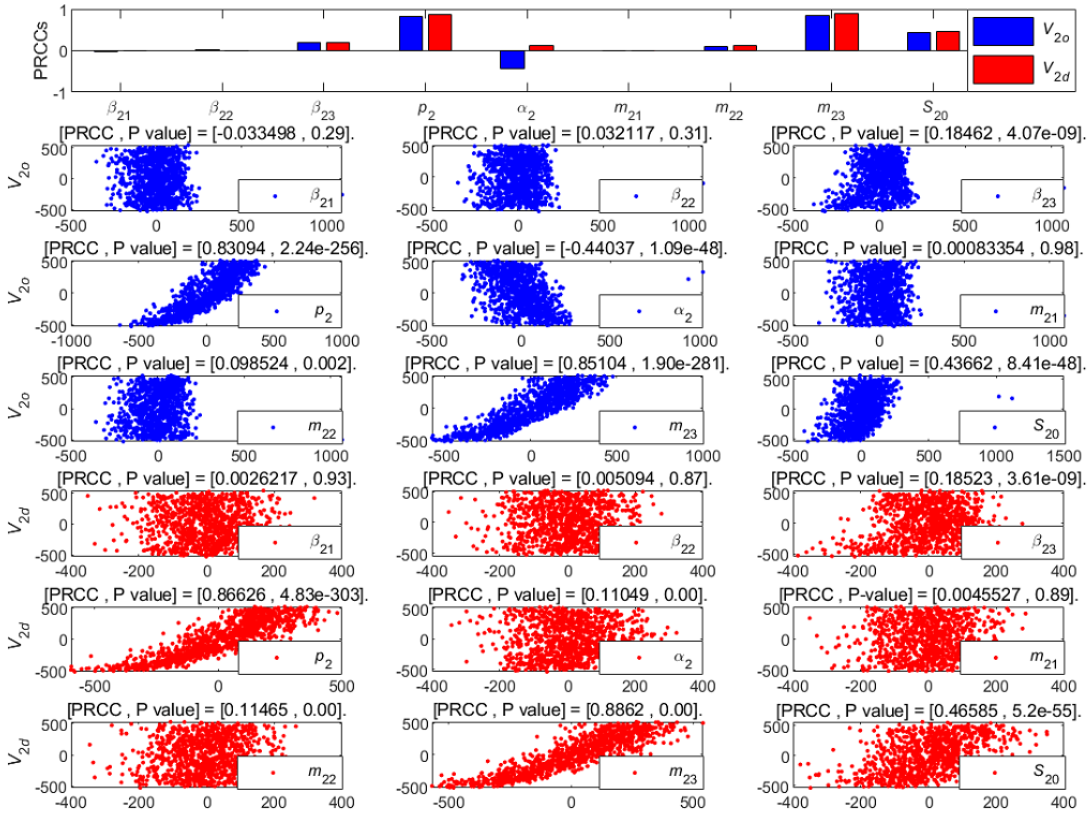


Figure 9 shows the effect of parameters on the public opinion reproduction ratio of the delay in transmission in the LTI DT-SFI model. was strongly positively affected by the average exposure rate β_{22}, the mild exposure attractiveness index m_{22}, the average exposure rate β_{23}, the unexposure attractiveness index m_{23}, and the forwarding probability p_2, and strongly negatively affected by the average immune rate α_2. The positive correlation effects of the parameters β_{21} and m_{21} were relatively weak. Overall, strategies to increase the parameters β_{22}, β_{23}, p_2, m_{22}, m_{23}, and initial value S_{20} or to decrease the α_2 can enhance the transmission capability of the newly posted information.

From Figure 10, the parameters have a similar effect on the forwarding peak F_{2max} and the cumulative forwarding population $C_{2\infty}$. The unexposure attractiveness index m_{23}, the forwarding probability p_2, and the initial value S_{20} of susceptible individuals have a decisive positive influence on the forwarding peak value F_{2max} and the final size $C_{2\infty}$ of delayed information propagation. The effects of the extensive and mild exposure attractiveness parameters that portray the participation of the population of the posted information were very weak. The aforementioned results indicate that the time interval is long between the two delays in transmission information since the new information was posted in a quasi–steady state into the propagation; at this time, most individuals who have been exposed to the posted information have entered the immune state. In addition, most individuals will no longer care about the relevant content due to the possibility of forgetting or leaving the social network platform. The aforementioned conclusions show that when the posted information enters the steady state, the effect of the individuals who have contacted the posted information is not obvious. Therefore, the information transmission can be promoted by influencing the number of the new susceptible population.

Figure 11 shows the effect of parameters on the climax time t_{2max}, the outbreak time t_{2b}, and the duration t_{2i} of the delay in transmission. After mastering the influencing factors of t_{2b} and t_{2i}, the end time of transmission t_{2e} can be calculated. The climax time t_{2max}, the outbreak time t_{2b}, and the duration t_{2i} are negatively affected by parameters β_{23}, p_2, and m_{23} in the same way. In comparison, these parameters have the least impact on t_{2b}, especially m_{23}. The parameter m_{22} had a weak negative correlation effect on each time index, and the parameter α_2 was the main factor to control the duration t_{2i}, which plays a strong negative correlation effect.

From Figure 12, the unexposure attractiveness index m_{23} and the forwarding probability p_2 had major positive correlation effects on the outbreak velocity V_{2o} and the declining velocity

V_{2d}, and the initial value S_{20} of susceptible individuals had a mild positive effect on these two indexes. Moreover, the parameter α_2 had a strong negative effect on the V_{2o}. In addition, the effects of other parameters on the velocities were not important. That is to say, the V_{2o} and V_{2d} will increase accordingly when the parameters m_{23}, p_2, and initial value S_{20} increase. At the same time, the V_{2o} increases with the reduction of parameter α_2. By contrast, the effect of m_{23} on the velocities was greater.

Our LTI DT-SFI model concentrates on the influence of the average exposure rates and attractiveness indexes on the instantaneous forwarding population $F_2(t)$ and the cumulative forwarding population $C_2(t)$ as shown in Figures 13 and 14, respectively, and the variation of parameters over time determines the propagation indexes. By comparing and analyzing the influence of average contact rates and attractiveness indexes in Figures 11 and 12 with the variation of one parameter while fixing other parameters, β_{23} and m_{23} have a similar overall trend of the effects on the instantaneous forwarding population $F_2(t)$ and the cumulative forwarding population $C_2(t)$ of the new information. With the increase of the parameters of β_{23} and m_{23}, the outbreak will accelerate, the instantaneous number of individuals in the forwarding state can reach a higher peak, and the final size will be larger. In addition, the average exposure rate of β_{22} and the mild exposure attractiveness index m_{22} had a weak positive influence on the final size of the cumulative forwarding quantity and had no obvious influence on the propagation times and velocities. In contrast, the average exposure rate β_{21} and the extensive exposure attractiveness index m_{21} had no significant effect on large interval delay in transmission based on forwarding. All the aforementioned key parameters had no significant effect on the outbreak time, climax time, and duration of the long-delayed cross-information transmission based on forwarding, which was also consistent with the results of the partial rank correlation coefficients.

Figure 13. The influence of the average exposure rates on the instantaneous forwarding population $F_2(t)$ and the cumulative forwarding population $C_2(t)$ in the large interval delay in transmission susceptible-forwarding-immune model.

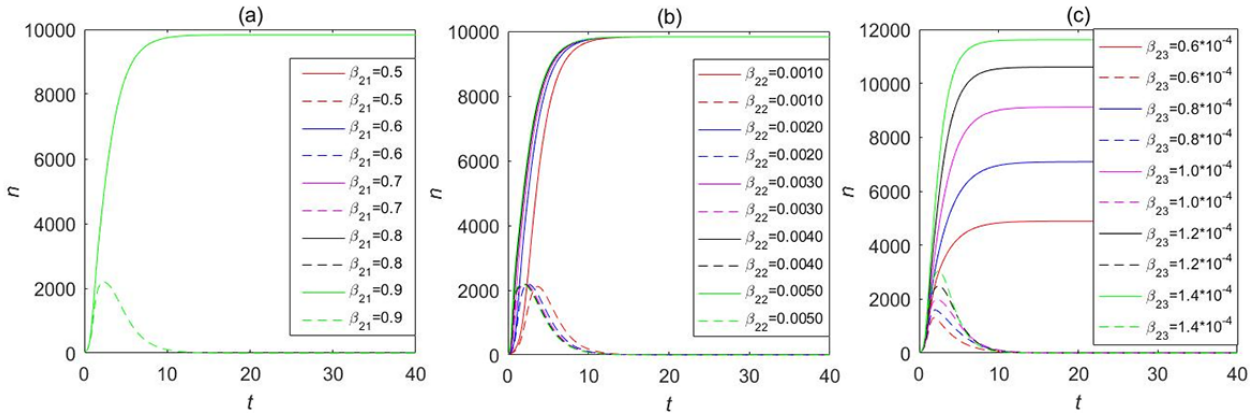
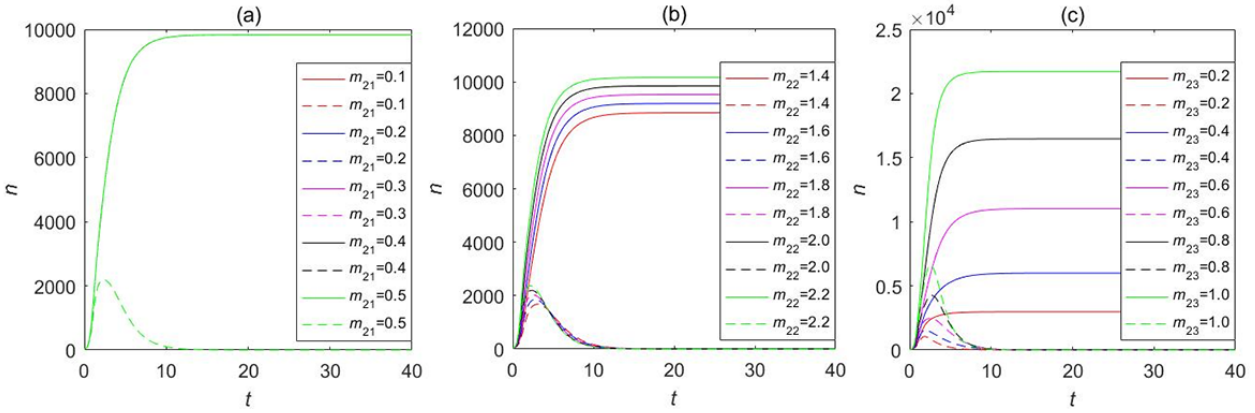


Figure 14. The influence of the attractiveness indexes on the instantaneous forwarding population $F_2(t)$ and the cumulative forwarding population $C_2(t)$ in the large interval delay in transmission susceptible-forwarding-immune model.



Influencing Factors Analytics: Information Release and Dissemination for the STI DT-SFI Model

To further analyze the impact of different parameters in the STI DT-SFI model for the cross-propagation dynamics, we performed partial rank correlation coefficients to analyze the relationship between the influence and the range of variation of parameters on the indexes. Figures 15-18 give the partial rank correlation coefficient results and partial rank correlation coefficient scatterplots with indexes , F_{2max}, $C_{2\infty}$, t_{2b}, t_{2i}, t_{2max}, V_{2o}, and V_{2d} with nine parameters (β_{21}, β_{22}, β_{23}, p_2, α_2, m_{21}, m_{22}, m_{23}, and S_{10}) of the newly posted information in the STI DT-SFI model, respectively.

Figure 15. PRCC results and PRCC scatterplots with indexes for different parameters of the newly posted information in the short interval delay in transmission–susceptible-forwarding-immune model. PRCC: partial rank correlation coefficient.

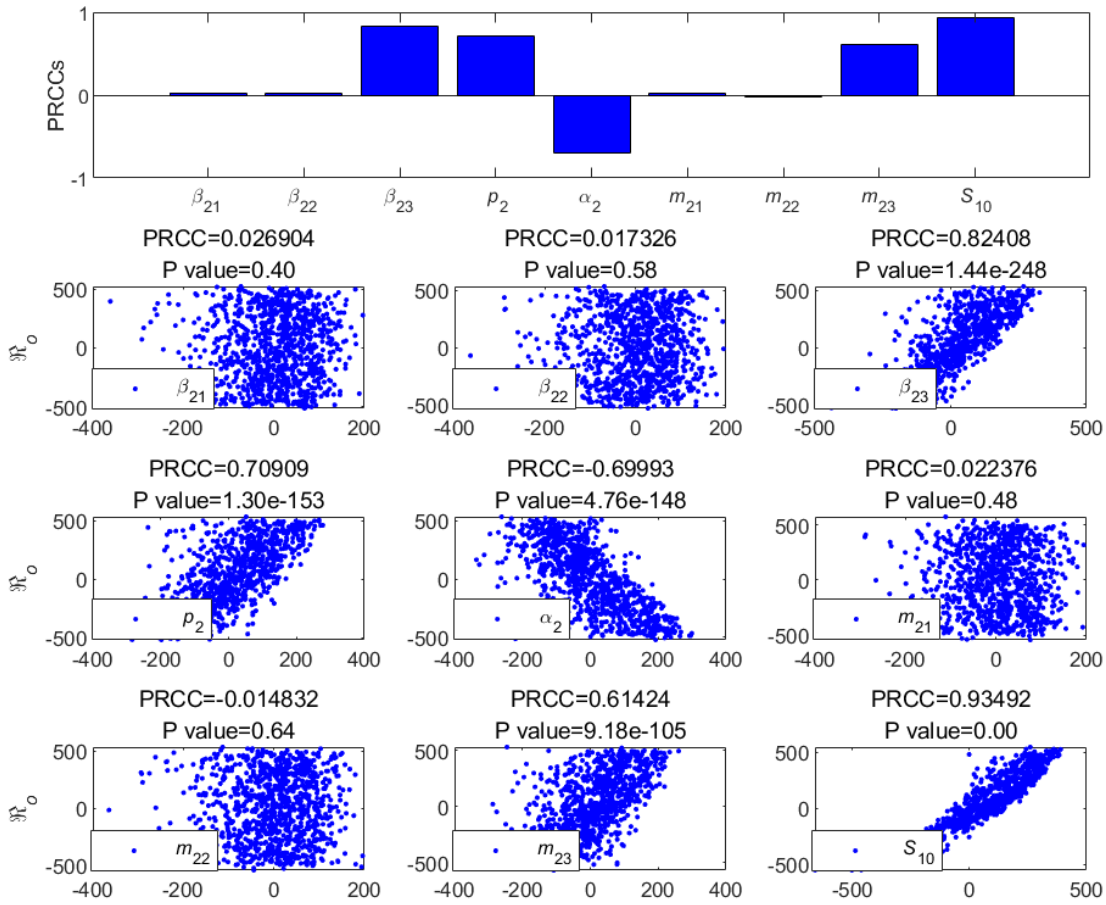


Figure 16. PRCC results and PRCC scatterplots with indexes F_{2max} and $C_{2\infty}$ for different parameters of the newly posted information in the short interval delay in transmission susceptible-forwarding-immune model. PRCC: partial rank correlation coefficient.

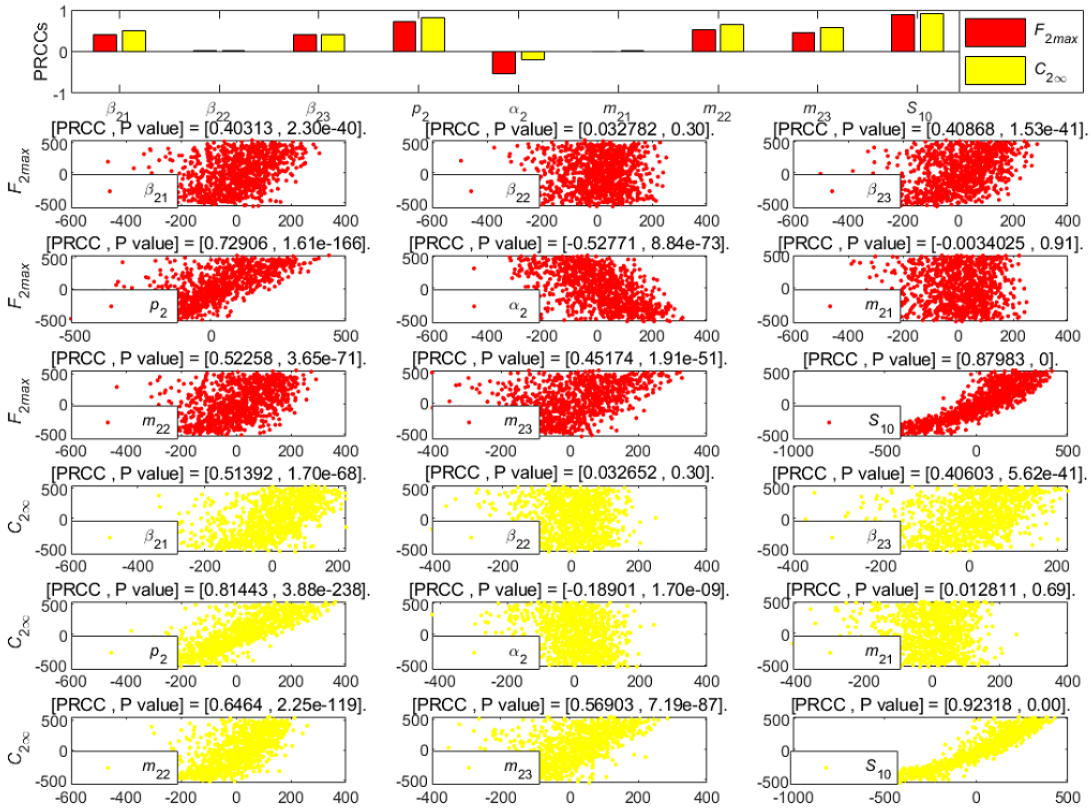


Figure 17. PRCC results and PRCC scatterplots with indexes t_{2max}, t_{2b}, and t_{2i} for different parameters of the newly posted information in the short interval delay in transmission susceptible-forwarding-immune model. PRCC: partial rank correlation coefficient.

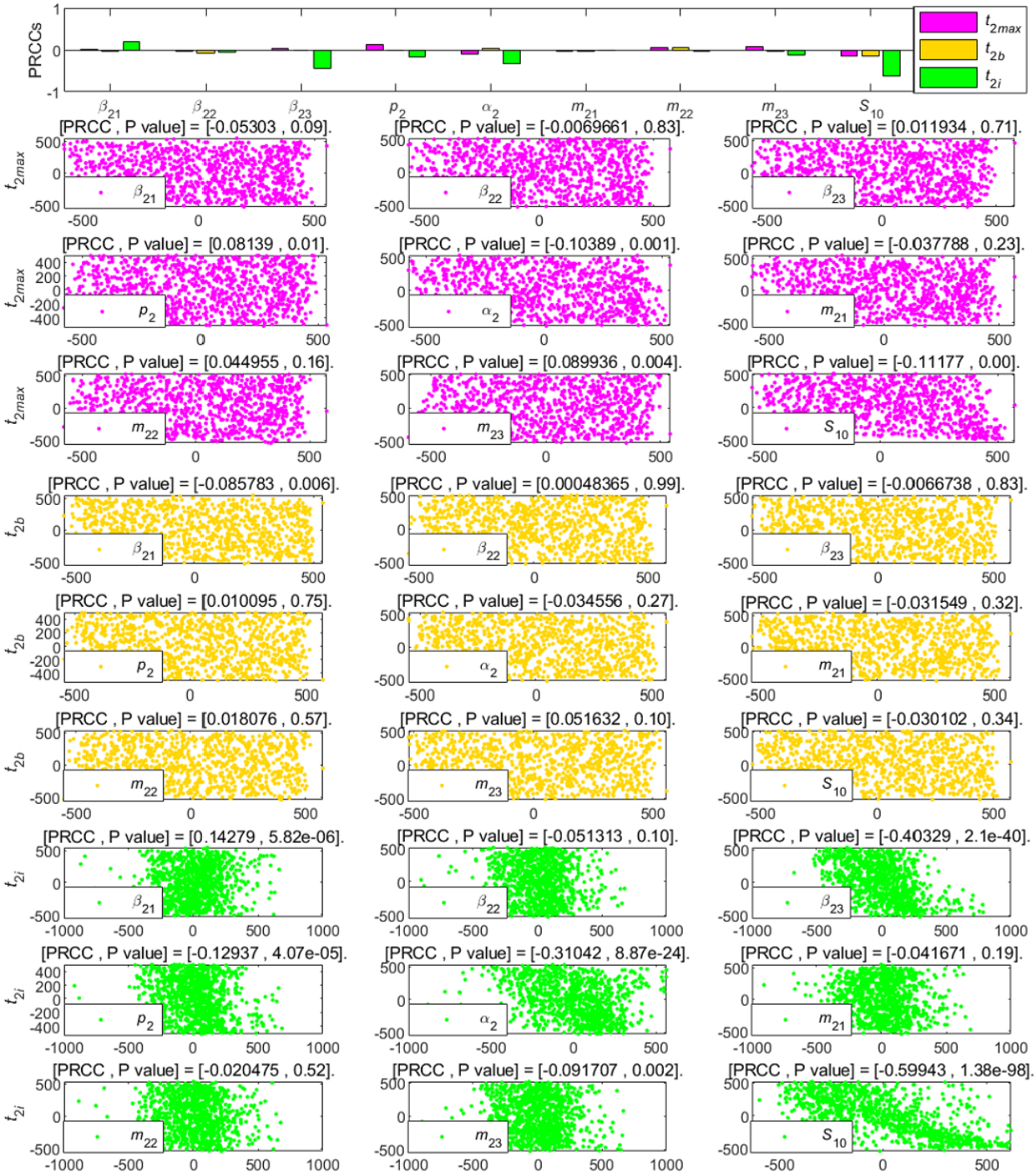
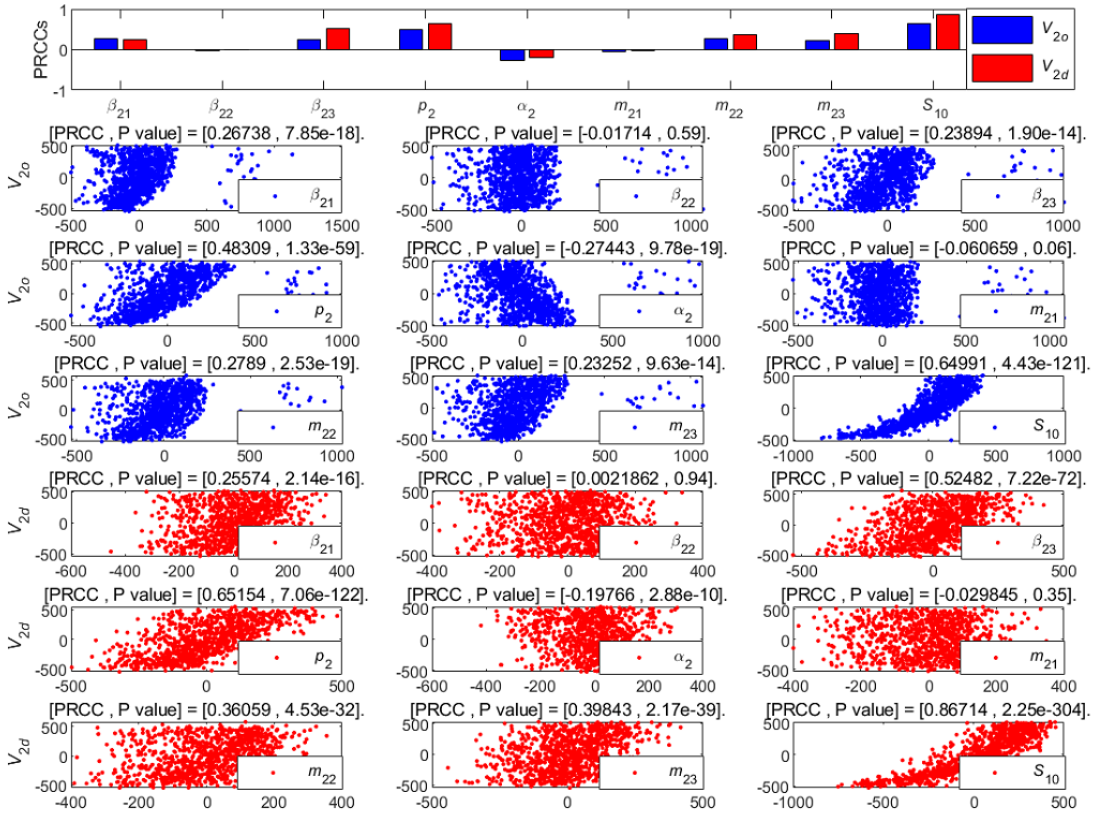


Figure 18. PRCC results and PRCC scatterplots with indexes V_{2o} and V_{2d} for different parameters of the newly posted information in the short interval delay in transmission susceptible-forwarding-immune model. PRCC: partial rank correlation coefficient.



The average exposure rate β_{23}, the unexposure attractiveness index m_{23}, the forwarding probability p_2, and the initial value S_{10} were strong contributions to the public opinion reproduction ratio positively, and the average immune rate α_2 had a strong negative effect on it, as shown in Figure 15. The positive correlation effects of parameters β_{21}, β_{22}, m_{21}, and m_{22} were relatively weak. In general, strategies that can affect the parameters β_{23}, m_{23}, p_2, and the initial value S_{10} to increase or the parameter α_2 to decrease can increase the initial propagation capacity of the newly posted information. On the other hand, we can decrease the parameters β_{23}, m_{23}, p_2, and the initial value S_{10} to reduce the initial propagation ability of the new information.

The average contact rate of β_{21}, β_{22}, the forwarding probability p_2, the mild exposure attractiveness index m_{22}, the unexposure attractiveness index m_{23}, and the initial value S_{10} of susceptible individuals had strong positive impacts on the high peak F_{2max} and the final size $C_{2\infty}$, as shown in Figure 16. In contrast, S_{10} played a major role, and the impact of β_{22} and m_{21} were less significant. The aforementioend results show that individuals who have been exposed to but have not forwarded the posted information are more sensitive to the new information with mild exposure attractiveness due to the understanding of the former information. In addition, the average immune rate α_2 had a strong negative effect on the F_{2max}.

Figure 17 indicates that the influence of each parameter on t_{2max} and t_{2b} were not obvious, and the parameters β_{23}, α_2, and the initial value S_{10} had negative effects on the duration t_{2i}, and the β_{21} had a weak positive effect on it. This means that the average contact rate at which users in the susceptible state can contact the second information is the most important factor affecting the duration t_{2i} of delay in transmission. The smaller the average contact rate is, the longer the duration of new information transmission will be within a certain range, slowing down the development of information transmission.

Figure 18 shows the partial rank correlation coefficients results of the outbreak velocity V_{2o} and the declining velocity V_{2d} of the STI DT-SFI model based on forwarding under multiparameter changes. From the results, the average exposure rate β_{21}, β_{23}, the forwarding probability p_2, the mild exposure attractiveness index m_{22}, the unexposure attraction index m_{23}, and the initial value S_{10} of susceptible individuals make strong positive contributions on V_{2o} and V_{2d}. The average exposure rate β_{22} and the extensive exposure attractiveness index m_{21} had no significant effect on the velocities. That is to say, the outbreak velocity V_{2o} and the declining velocity V_{2d} can be increased with the increase of the parameters β_{21}, β_{23}, p_2, m_{22}, m_{23}, and

the initial value S_{10}. On the contrary, if the parameters decrease, the propagation velocities will slow down.

Here, we also took into consideration the influence of the average exposure rate and attractiveness parameters on the instantaneous forwarding population $F_2(t)$ and the cumulative forwarding population $C_2(t)$ of the STI DT-SFI model as shown in Figures 19 and 20, respectively. The comparative analysis shows that the larger the average contact rate and attractiveness indexes are, the larger the instantaneous forwarding quantity and the cumulative forwarding quantity are. The final size is also affected; the average exposure rate of β_{21}, β_{23}, the mild exposure attractiveness index m_{22}, and the unexposure attractiveness index m_{23} are the main influencing factors of the STI DT-SFI model, and they can play a significant role in the final size of the newly posted information within a certain range. So priority must be placed on controlling these parameters. In addition, the extensive exposure attractiveness index m_{21} has only a small magnitude of effects, while the effect of parameter β_{22} is significant and has a relatively obvious impact. The impact of each parameter on the outbreak timing and increasing and declining velocities is negligible, which is consistent with the results of the partial rank correlation coefficients.

Figure 19. The influence of the average exposure rates on the instantaneous forwarding population $F_2(t)$ and the cumulative forwarding population $C_2(t)$ in the short interval delay in transmission susceptible-forwarding-immune model.

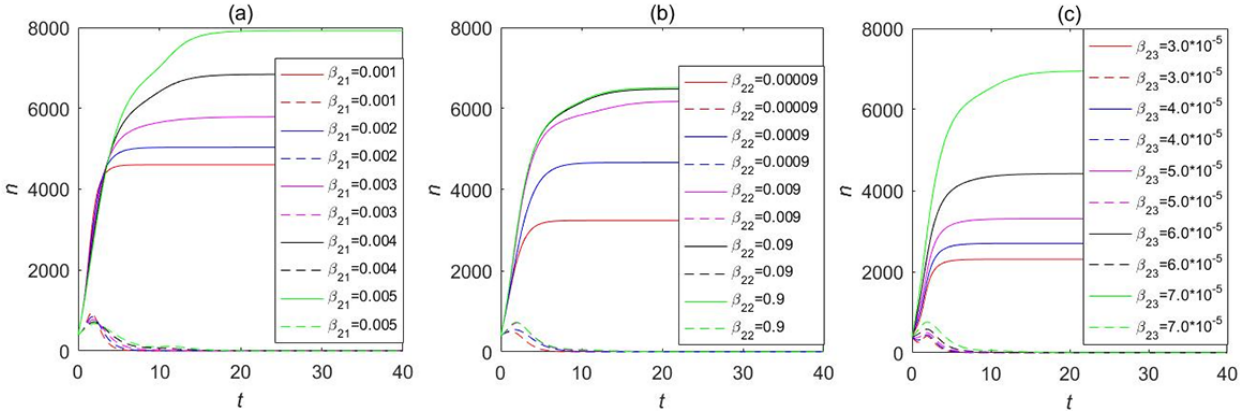
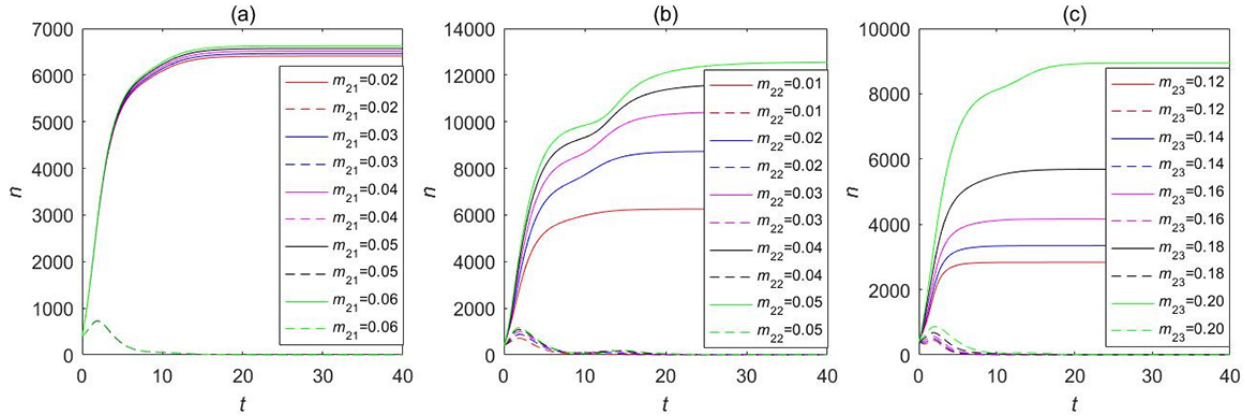


Figure 20. The influence of the attractiveness parameters on the instantaneous forwarding population $F_2(t)$ and the cumulative forwarding population $C_2(t)$ in the short interval delay in transmission susceptible-forwarding-immune model.



Discussion

Principal Findings

Figure 21 shows the trend of the cumulative forwarding users of information B and information C when the newly posted information is posted during the steady-state period of the posted information in Tables 2 and 3. The time lag with which the newly posted information is posted has a significant impact on the process of public opinion dissemination and the final size of the cumulative number of forwarding users. If the newly posted information is posted during the quasi–steady-state period of opinion propagation, then the earlier the newly posted information is posted, the earlier the cumulative forwarding users will peak, though the final size of the cumulative forwarding users will be close to each other. This, in conjunction with our parameter sensitivity analysis results, shows that for the LTI DT-SFI situation, the unexposure attractiveness index m_{23} and the average exposure rate β_{23} are the key elements to promote the cross-propagation and that, once reaching the quasi–steady state, the timing of posting the new information has an insignificant impact on the final size of forwarding users.

In contrast, Figure 22 shows the trend of the cumulative forwarding users of information A and information B when the newly posted information is posted during the outbreak period of the posted information in Tables 1 and 2. The time lag with which the newly posted information is posted has a noticeable impact on both the dynamic process of public opinion dissemination and the final size of the cumulative number of

forwarding users. If the newly posted information can be posted during the outbreak period of the old information, then the earlier the new information is posted, the greater the cumulative forwarding population and the final size of the cumulative forwarding users will be. This, combined with our parameter sensitivity analysis results, shows that, in the STI DT-SFI case, the mild exposure attractiveness index m_{22}, the average exposure rates β_{21} and β_{23}, and the unexposure attractiveness index m_{23} can all directly influence the interaction between information posted sequentially to increase the "heat" (popularity) of the newly posted information.

Our model-based analysis recommends strategies on how different parameters should be adjusted to achieve the best information dissemination outcomes. For two pieces of relevant information separated by a relative long posting lag, strategies to increase the average exposure rate β_{23} and the unexposure attractiveness index m_{23} are recommendations. These strategies can be achieved if opinion leaders with a large number of followers can participate in the information copropagation. On the contrary, reducing the public's attention to a new piece of information can be achieved by efforts in delaying the posting of the new information or by effectively reducing the potential correlation between the two pieces of information (reducing values of the correlation parameters β_{21}, β_{22}, m_{21}, and m_{22}). Additionally, if our goal is for the final size of the cumulative forwarding users of the new information to not be impacted by the relevant information already posted online, the new information should be posted during the quasi–steady-state period of the posted information.

For two pieces of information with a short interval between posting, we recommend developing strategies to alter the interaction between the information for effectively managing the information transmission indexes we introduced. If we aim to make the new information outbreak faster with a large peak value of forwarding, we should increase the average exposure rate β_{21} and mild exposure attractiveness index m_{22} by persuading the original post owner to post or forward the information earlier during the outbreak period of the posted information, when the posted information has obtained certain public attention, and increase the relevance and attraction of the newly posted information to the forwarding users or immune users of the posted information. Alternatively, we should persuade some opinion leaders to forward the new information along with their insights to β_{23} and m_{23}.

Figure 21. An illustration of a public opinion dissemination process with newly posted information posted with different time lags but during the quasi–steady-state period of the posted information.

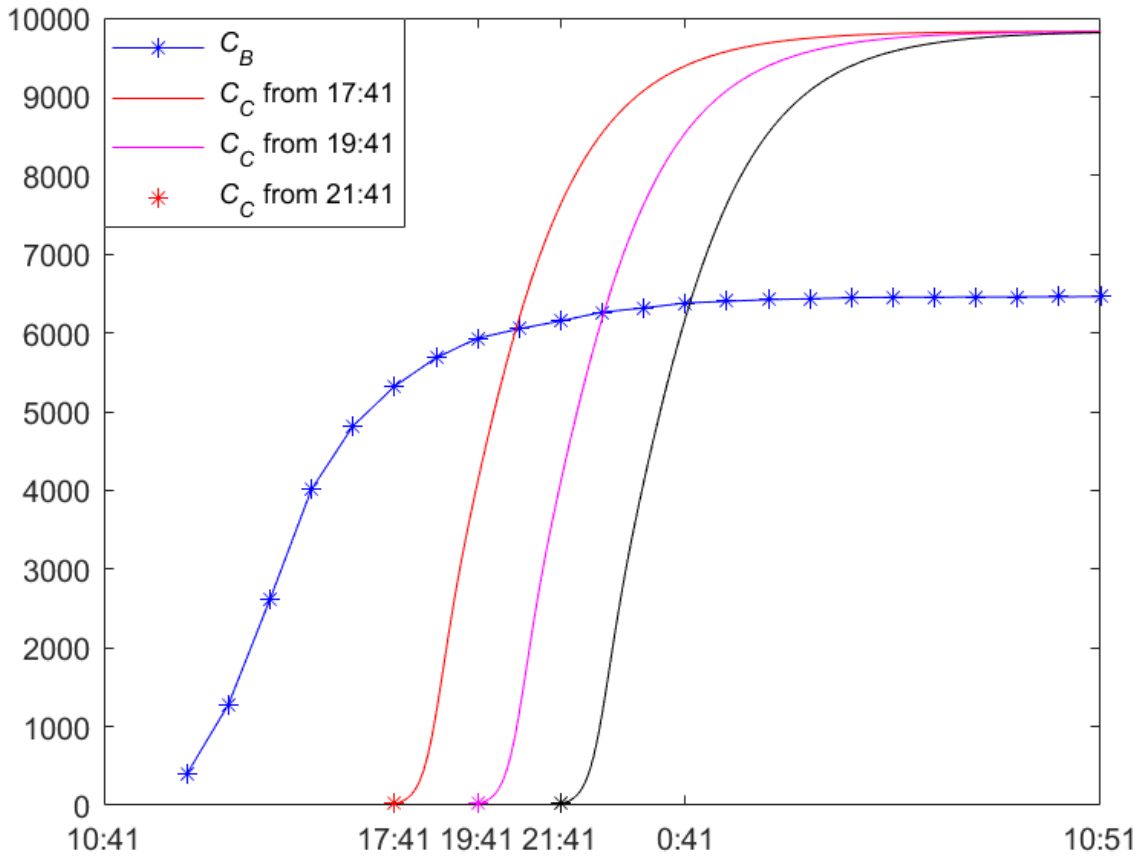
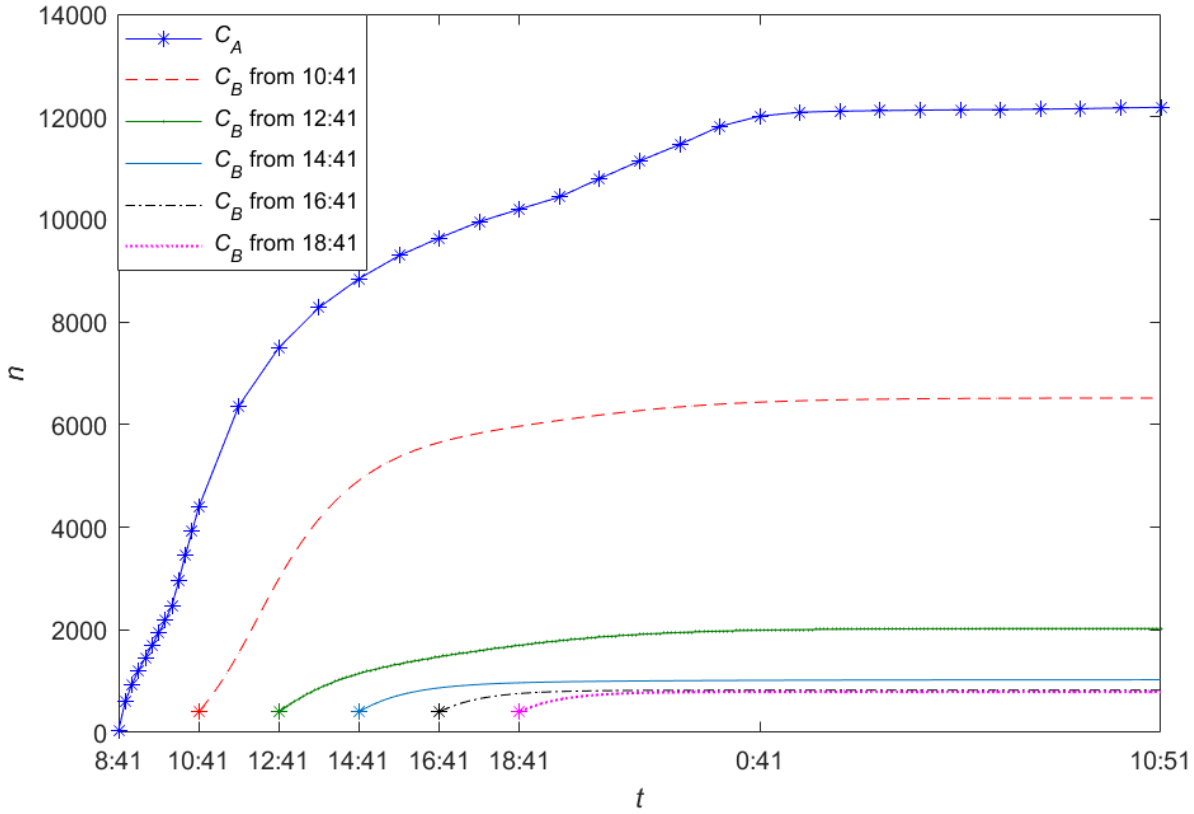


Figure 22. An illustration of a public opinion dissemination process with newly posted information posted with different time lags but during the quasi–steady-state period of the posted information.



Conclusions

Relevant information is often posted sequentially in fast-evolving public health events such as the COVID-19 pandemic. Consequently, modeling the impact of delay in cross-transmission or copropagation is substantial to identifying the best strategies to communicate key public messages through social media. In this study, we proposed and examined two classes of models, the LTI DT-SFI dynamics model and the STI DT-SFI dynamics model based on the forwarding users in Weibos, and we parametrized our models using real data related to the COVID-19 pandemic in the Chinese Sina Microblog. Our goal is to use these parametrized models to understand the influence of different time lags in the information posting on the copropagation of related information in the microblog.

Our model formulation focused on the transmission mechanism of information in the social network, where a new Weibo may be posted in different phases—outbreak phase or quasi–steady-state phase—of some relevant Weibo already posted. Our goal is to examine the impact of post timing in relation to the old information, the new information on its peak value, and the final size of forwarding users. As we have shown, this impact depends on the correlation of the old and new information, and on the phase of the old information transmission when the new information is posted. We hope that our DT-SFI dynamics models fill in some theoretical gap about optimizing information posting strategies to maximize communication efforts to deliver key public health messages to the public for better outcomes of public health emergency management.

Acknowledgments

The study was supported by the National Natural Science Foundation of China (Grant number: 61801440), the Natural Science and Engineering Research Council of Canada, the Canada Research Chair Program (JW), the Fundamental Research Funds for the Central Universities and the High-quality and Cutting-edge Disciplines Construction Project for Universities in Beijing (Internet Information, Communication University of China), the State Key Laboratory of Media Convergence and Communication, and the Communication University of China.

Conflicts of Interest

None declared.

References

1. Goffman W, Newill VA. Generalization of epidemic theory. An application to the transmission of ideas. Nature 1964 Oct 17;204:225-228. [doi: 10.1038/204225a0] [Medline: 14212412]
2. Xiao D, Ruan S. Global analysis of an epidemic model with nonmonotone incidence rate. Math Biosci 2007 Aug;208(2):419-429 [FREE Full text] [doi: 10.1016/j.mbs.2006.09.025] [Medline: 17303186]
3. Li MY, Muldowney JS. Global stability for the SEIR model in epidemiology. Math Biosci 1995 Feb;125(2):155-164. [doi: 10.1016/0025-5564(95)92756-5] [Medline: 7881192]
4. Biswas MHA, de Pinho MDR. A SEIR model for control of infectious diseases with constraint. Math Biosciences Eng 2014;11(4):761-784. [doi: 10.3934/mbe.2014.11.761]
5. Li MY, Smith H, Wang L. Global dynamics of an SEIR epidemic model with vertical transmission. SIAM J Appl Math 2001 Jan;62(1):58-69. [doi: 10.1137/S0036139999359860]
6. Chen FH. A susceptible-infected epidemic model with voluntary vaccinations. J Math Biol 2006 Aug;53(2):253-272. [doi: 10.1007/s00285-006-0006-1] [Medline: 16758209]
7. Lu Z, Gao S, Chen L. Analysis of an SI epidemic model with nonlinear transmission and stage structure. Acta Mathematica Scientia 2003 Oct;23(4):440-446. [doi: 10.1016/s0252-9602(17)30486-1]
8. Zhang T, Meng X, Song Y, Zhang T. A stage-structured predator-prey SI model with disease in the prey and impulsive effects. Math Modelling Analysis 2013;18(4):505-528.
9. Yuan C, Hu B. A SI epidemic model with two-stage structure. Mathematicae Applicatae Sinica 2002;25(2):193-203.
10. Xia C, Sun S, Rao F, Sun J, Wang J, Chen Z. SIS model of epidemic spreading on dynamical networks with community. Front Comput Sci China 2009 Aug 15;3(3):361-365. [doi: 10.1007/s11704-009-0057-8]
11. Shi H, Duan Z, Chen G. An SIS model with infective medium on complex networks. Physica A Stat Mechanics Applications 2008 Mar;387(8-9):2133-2144. [doi: 10.1016/j.physa.2007.11.048]
12. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics--I. 1927. Bull Math Biol 1991;53(1-2):33-55. [doi: 10.1007/BF02464423] [Medline: 2059741]
13. Stone L, Shulgin B, Agur Z. Theoretical examination of the pulse vaccination policy in the SIR epidemic model. Math Comput Modelling 2000 Feb;31(4-5):207-215. [doi: 10.1016/s0895-7177(00)00040-6]
14. Xia C, Wang L, Sun S, Wang J. An SIR model with infection delay and propagation vector in complex networks. Nonlinear Dyn 2012 Jan 6;69(3):927-934. [doi: 10.1007/s11071-011-0313-y]
15. Gu J, Cai X. The forget-remember mechanism for 2-state spreading. arXiv. Preprint posted online February 11, 2007.
16. Gu J, Li W, Cai X. The effect of the forget-remember mechanism on spreading. Eur Phys J B 2008 Apr 4;62(2):247-255. [doi: 10.1140/epjb/e2008-00139-4]
17. Zhao L, Wang Q, Cheng J, Chen Y, Wang J, Huang W. Rumor spreading model with consideration of forgetting mechanism: A case of online blogging LiveJournal. Physica A Stat Mechanics Applications 2011 Jul;390(13):2619-2625. [doi: 10.1016/j.physa.2011.03.010]
18. Zhao L, Wang X, Wang J, Qiu X, Xie W. Rumor-propagation model with consideration of refutation mechanism in homogeneous social networks. Discrete Dynamics Nat Soc 2014;2014:1-11. [doi: 10.1155/2014/659273]
19. Huang J, Su Q. A rumor spreading model based on user browsing behavior analysis in microblog. 2013 Presented at: 10th International Conference on Service Systems and Service Management; July 17-19, 2013; Hong Kong, China p. 170-173. [doi: 10.1109/icsssm.2013.6602630]
20. Zhao L, Cui H, Qiu X, Wang X, Wang J. SIR rumor spreading model in the new media age. Physica A Stat Mechanics Applications 2013 Feb;392(4):995-1003. [doi: 10.1016/j.physa.2012.09.030]
21. Borge-Holthoefer J, Meloni S, Gonçalves B, Moreno Y. Emergence of influential spreaders in modified rumor models. J Stat Phys 2012 Sep 22;151(1-2):383-393. [doi: 10.1007/s10955-012-0595-6]
22. Yin F, Lv J, Zhang X, Xia XY, Wu JH. COVID-19 information propagation dynamics in the Chinese Sina-microblog. Math Biosci Eng 2020 Mar 09;17(3):2676-2692 [FREE Full text] [doi: 10.3934/mbe.2020146] [Medline: 32233560]
23. Ding X, Liu Q, Zhang W. An improved model for information dissemination and prediction on microblog networks. J University Sci Technol China 2014;44(7):582.
24. Wang H, Li Y, Feng Z, Feng L. ReTweeting analysis and prediction in microblogs: An epidemic inspired approach. China Commun 2013 Mar;10(3):13-24. [doi: 10.1109/cc.2013.6488827]
25. Zhang Y, Feng Y, Yang R. Network public opinion propagation model based on the influence of media and interpersonal communication. Int J Modern Phys B 2020 Jan 10;33(32):1950393. [doi: 10.1142/s0217979219503934]
26. Zhao L, Wang J, Chen Y, Wang Q, Cheng J, Cui H. SIHR rumor spreading model in social networks. Physica A Stat Mechanics Applications 2012 Apr;391(7):2444-2453. [doi: 10.1016/j.physa.2011.12.008]
27. Woo J, Ha SH, Chen H. Tracing topic discussions with the event-driven SIR model for online forums. J E Commerce Res 2016;17(2):169.
28. Yao Y, Li Y, Xiong X, Wu Y, Lin H, Ju S. An interactive propagation model of multiple information in complex networks. Physica A Stat Mechanics Applications 2020 Jan;537:122764. [doi: 10.1016/j.physa.2019.122764]
29. Trpevski D, Tang WKS, Kocarev L. Model for rumor spreading over networks. Phys Rev E Stat Nonlin Soft Matter Phys 2010 May;81(5 Pt 2):056102. [doi: 10.1103/PhysRevE.81.056102] [Medline: 20866292]

30. Qian Z, Tang S, Zhang X, Zheng Z. The independent spreaders involved SIR Rumor model in complex networks. Physica A Stat Mechanics Applications 2015 Jul;429:95-102. [doi: 10.1016/j.physa.2015.02.022]
31. Wang J, Zhao L, Huang R. SIRaRu rumor spreading model in complex networks. Physica A Stat Mechanics Applications 2014 Mar;398:43-55. [doi: 10.1016/j.physa.2013.12.004]
32. Tanaka M, Sakumoto Y, Aida M, Kawashima K. Study on the growth and decline of SNSs by using the infectious recovery SIR model. 2015 Presented at: 10th Asia-Pacific Symposium on Information and Telecommunication Technologies (APSITT); August 4-7, 2015; Colombo, Sri Lanka p. 1-3. [doi: 10.1109/apsitt.2015.7217089]
33. Feng X, Huo X, Tang B, Tang S, Wang K, Wu J. Modelling and analyzing virus mutation dynamics of Chikungunya outbreaks. Sci Rep 2019 Feb 27;9(1):2860. [doi: 10.1038/s41598-019-38792-4] [Medline: 30814598]
34. Zan Y. DSIR double-rumors spreading model in complex networks. Chaos Solitons Fractals 2018 May;110:191-202. [doi: 10.1016/j.chaos.2018.03.021]
35. Yin F, Shao X, Wu J. Nearcasting forwarding behaviors and information propagation in Chinese Sina-Microblog. Math Biosci Eng 2019 Jun 11;16(5):5380-5394 [FREE Full text] [doi: 10.3934/mbe.2019268] [Medline: 31499717]
36. Abbas UL, Anderson RM, Mellors JW. Potential impact of antiretroviral chemoprophylaxis on HIV-1 transmission in resource-limited settings. PLoS One 2007 Sep 19;2(9):e875 [FREE Full text] [doi: 10.1371/journal.pone.0000875] [Medline: 17878928]

Abbreviations

DT-SFI: delay in transmission susceptible-forwarding-immune
LS: least squares
LTI DT-SFI: large interval delay in transmission susceptible-forwarding-immune
SIR: susceptible-infected-recovered
STI DT-SFI: short interval delay in transmission susceptible-forwarding-immune

Edited by R Kukafka; submitted 17.11.20; peer-reviewed by H Wang, X Feng; comments to author 01.12.20; revised version received 29.12.20; accepted 31.01.21; published 12.02.21.

Please cite as:
Yin F, Shao X, Ji M, Wu J
Quantifying the Influence of Delay in Opinion Transmission of COVID-19 Information Propagation: Modeling Study
J Med Internet Res 2021;23(2):e25734
URL: http://www.jmir.org/2021/2/e25734/
doi:10.2196/25734
PMID:33529153

Original Paper

Dissemination and Refutation of Rumors During the COVID-19 Outbreak in China: Infodemiology Study

Bin Chen[1,2*], MD; Xinyi Chen[3*], MD; Jin Pan[4*], MD; Kui Liu[1*], MD; Bo Xie[5], PhD; Wei Wang[1], MD; Ying Peng[1], MD; Fei Wang[1], MD; Na Li[6], MD; Jianmin Jiang[1,7], MD, PhD

[1]Department of Tuberculosis Control and Prevention, Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China

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

[3]School of Medicine, Department of Preventative Medicine, Ningbo University, Ningbo, China

[4]Department of Non-communicable Disease Prevention, Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China

[5]School of Urban Design, Wuhan University, Wuhan, China

[6]Department of Public Health Emergency Response, Zhejiang Provincial Center for Disease Control and Prevention, Hangzhou, China

[7]Key Laboratory of Vaccine, Prevention and Control of Infectious Disease of Zhejiang Province, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Jianmin Jiang, MD, PhD
Department of Tuberculosis Control and Prevention
Zhejiang Provincial Center for Disease Control and Prevention
Binsheng Road 3399
Binjiang District
Hangzhou, 310051
China
Phone: 86 571 87115009
Fax: 86 571 87115009
Email: jmjiang@cdc.zj.cn

Abstract

Background: During the outbreak of COVID-19, numerous rumors emerged on the internet in China and caused confusion among the public. However, the characteristics of these rumors in different phases of the epidemic have not been studied in depth, and the official responses to the rumors have not been systematically evaluated.

Objective: The aims of this study were to evaluate the rumor epidemic and official responses during the COVID-19 outbreak in China and to provide a scientific basis for effective information communication in future public health crises.

Methods: Data on internet rumors related to COVID-19 were collected via the Sina Weibo Official Account to Refute Rumors between January 20 and April 8, 2020, extracted, and analyzed. The data were divided into five periods according to the key events and disease epidemic. Different classifications of rumors were described and compared over the five periods. The trends of the epidemic and the focus of the public at different stages were plotted, and correlation analysis between the number of rumors and the number of COVID-19 cases was performed. The geographic distributions of the sources and refuters of the rumors were graphed, and analyses of the most frequently appearing words in the rumors were applied to reveal hotspots of the rumors.

Results: A total of 1943 rumors were retrieved. The median of the response interval between publication and debunking of the rumors was 1 day (IQR 1-2). Rumors in text format accounted for the majority of the 1943 rumors (n=1241, 63.9%); chat tools, particularly WeChat (n=1386, 71.3%), were the most common platform for initial publishing of the rumors (n=1412, 72.7%). In addition to text rumors, Weibo and web pages were more likely to be platforms for rumors released in multimedia formats or in a combination of formats, respectively. Local agencies played a large role in dispelling rumors among social media platforms (1537/1943, 79.1%). There were significant differences in the formats and origins of rumors over the five periods (P<.001). Hubei Province accounted for most of the country's confirmed rumors. Beijing and Wuhan City were the main centers for debunking of disinformation. The words most frequently included in the core messages of the rumors varied by period, indicating shifting in the public's concern.

Conclusions: Chat tools, particularly WeChat, became the major sources of rumors during the COVID-19 outbreak in China, indicating a requirement to establish rumor monitoring and refuting mechanisms on these platforms. Moreover, targeted policy adjustments and timely release of official information are needed in different phases of the outbreak.

(J Med Internet Res 2021;23(2):e22427) doi:10.2196/22427

KEYWORDS

rumor; Internet; COVID-19; epidemic; misinformation; infodemiology; infodemic; social media; communication; public health

Introduction

In December 2019, an outbreak of COVID-19, caused by infection with SARS-CoV-2, emerged in Wuhan, Hubei Province, China, and subsequently became a global pandemic [1-3]. As of July 1, 2020, more than 10.3 million cases had been confirmed in most countries and territories worldwide, with more than 508,000 fatalities [4], seriously threatening the lives and health of the public and jeopardizing the stable economic development and social safety of nations.

Internet public opinion crises characterized by frequent rumors often accompany public health emergencies, especially when information from official authorities is delayed or lacking [5]. Since the advent of Web 2.0 technology, internet social media platforms such as WeChat (similar to WhatsApp) and Weibo (similar to Twitter) have gradually replaced traditional media as the main platforms for the public to express their opinion and participate in social affairs in China. Due to the easy accessibility and convenience of social media, information spreads more rapidly and widely through these platforms than through their conventional counterparts [6]; moreover, the resulting large availability of user-provided content fosters massive recruitment of people around common interests, worldviews, and narratives, thus affecting the evolution of public opinion [7] and further enabling rumors to flourish. In 2013, the World Economic Forum described web-based rumors as “digital wildfire” and highlighted the risks they pose to modern society [8]. The rumor that drinking spirit can prevent infection with SARS-CoV-2 is a typical example of a rumor that spread during the outbreak of COVID-19 in China. From the beginning of the outbreak, social media users started to query about methods of preventing and treating COVID-19, and they rushed to the internet to seek information. Due to strong concerns about their own lives and the lack of awareness of the disease, many microbloggers released messages that misrepresented the causal relationship between COVID-19 and drinking spirit, and their posts became very prevalent [9]. This false message was widely discussed on the internet at the time and caused great confusion and panic. As a result, the government provided an official clarification of the rumor, and various localities promptly refuted the rumor and addressed it through the intervention of public security departments, which prevented the rumor from spreading further.

According to previous research, public opinion events are often caused by the interaction of events, the public, social media platforms, and structural factors of the government [10,11]. Public health emergencies, especially outbreaks of new infectious diseases, are often accompanied by uncertainty about the cause of the emergency. However, the resulting information on the morbidity and mortality of the diseases becomes the focus of public concern from the moment it emerges. Individuals’ perceptions of the threat of diseases tends to be reinforced by their exposure to case data and also by public and private information that is disseminated widely on social media [12]. Simultaneously, the unknown causes of public health emergencies stimulate increased information-seeking behavior in people who are aiming to reduce their uncertainties about the emergent situation [13]. However, in the absence of information, people experience a wide range of emotions in the face of unexpected situations, and the anxiety or fear thus generated can exacerbate the occurrence and dissemination of rumors [14]. The role of government intervention in the development trend of rumors remains uncertain. However, the subject, duration, methods, and level of government intervention have certain influences on the virality of rumors [10]. In addition to the professional measures of epidemic prevention and control, keeping the information accurate and transparent and preventing the spread of rumors are critical parts of the crisis response, reflecting the significance of the establishment of government monitoring-feedback-intervention mechanisms in public health emergencies [15].

Compared with the severe acute respiratory syndrome (SARS) outbreak 17 years ago, the COVID-19 outbreak has sparked more rumormongering. Rumors such as “dual yellow oral liquid inhibited novel coronavirus,” “number of confirmed cases of COVID-19 and deaths in a county,” and “some places have been blockaded or the supermarkets have been closed down” sparked panic among the public, causing people to rush to buy supplies and posing a serious challenge to the governance of internet public opinion in the context of the epidemic. As the challenges grew in the face of the public crisis, the phenomenon of the “infodemic,” an overabundance of accurate or inaccurate information occurring during an epidemic, has escalated to a level that requires a coordinated response. Thus, the emerging research area of “infodemiology,” which can be defined as the science of using epidemiologic methods and terminology to study the distribution and determinants of information in an electronic medium, specifically the internet, with the ultimate aim to inform public health and public policy, has been developed [16] and was effectively used to predict the influenza outbreak in 2006 [17]. Infodemiology data are derived from unstructured, textual, openly accessible information produced and consumed by the public on the internet to demonstrate and explore the opinions, focus, behavior, attitudes, and knowledge of the public [16]. The research field of infodemiology has gradually gained wider use, and it caught the attention of the World Health Organization in the wake of the COVID-19 outbreak [18], encouraging the undertaking of more relevant

studies and effective practices to understand more about internet information.

Thus, in this study, we analyzed rumors collected from rumor-refuting platforms, using the methods of infodemiology from the supply side, to understand the epidemic of rumors and official responses in different periods according to typical events and the disease epidemic. The results of this study could provide evidence-based recommendations for information communication and rumor prevention during subsequent public health emergencies.

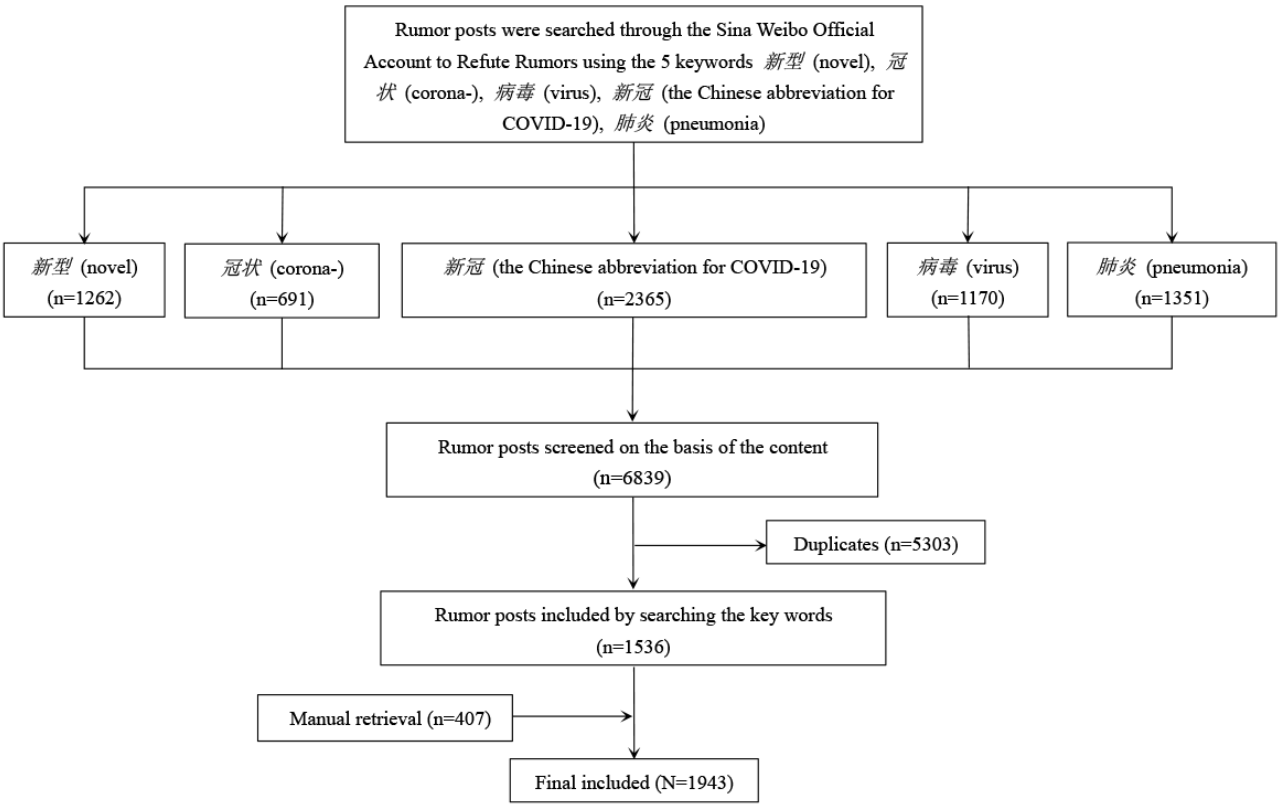
Methods

Data Sources

Between January 20, 2020 (the day the national authority first published the official announcement of human-to-human transmission of COVID-19), and April 8, 2020 (the day the lockdown was lifted in Wuhan), daily reports of identified and confirmed rumors and counterrumors of the COVID-19 outbreak were collected through the Sina Weibo Official Account to Refute Rumors (hereinafter referred to as "the Rumors on Weibo account") [19]. Launched in 2010, this account has been dedicated to rumor-busting and has now become one of the largest accounts in China, with more than 2.16 million registered users and subscribers; it contains a massive amount of information collected from most official platforms, such as the Chinese National Platform to Refute Rumors, WeChat Disinformation Platform, internet media, and web pages of government authorities, to refute rumors. Each post in the Rumors on Weibo account contains a rumor message and related rumor-refuting information.

During the first week of the study period (January 20-27), a small sample of 311 rumor messages was collected via the Rumors on Weibo account, and word frequency analysis was performed on these messages. After a panel discussion, 新型 (novel), 冠状 (corona-), 病毒 (virus), 新冠 (the Chinese abbreviation for COVID-19), and 肺炎 (pneumonia) were selected as the 5 keywords according to their frequency in the rumor posts. A total of 6839 disinformation messages were retrieved on the Rumors on Weibo account based on these keywords during the whole study period, of which 5303 were duplicated; eventually, 1536 rumor messages related to the epidemic were included. Meanwhile, other rumor posts not containing the above keywords were reviewed manually in the Rumors on Weibo account each week. A total of 407 relevant rumor messages, not containing the above 5 keywords, were also retrieved and added to the database (Figure 1). The initial posting time, title, posting platform, geographic location, format of each rumor, rumor-refuting time, and number of retweets of each rumor post were extracted and entered into the database. All the information obtained on the web was in simplified Chinese language and released publicly by the websites; however, no personal identification information was collected.

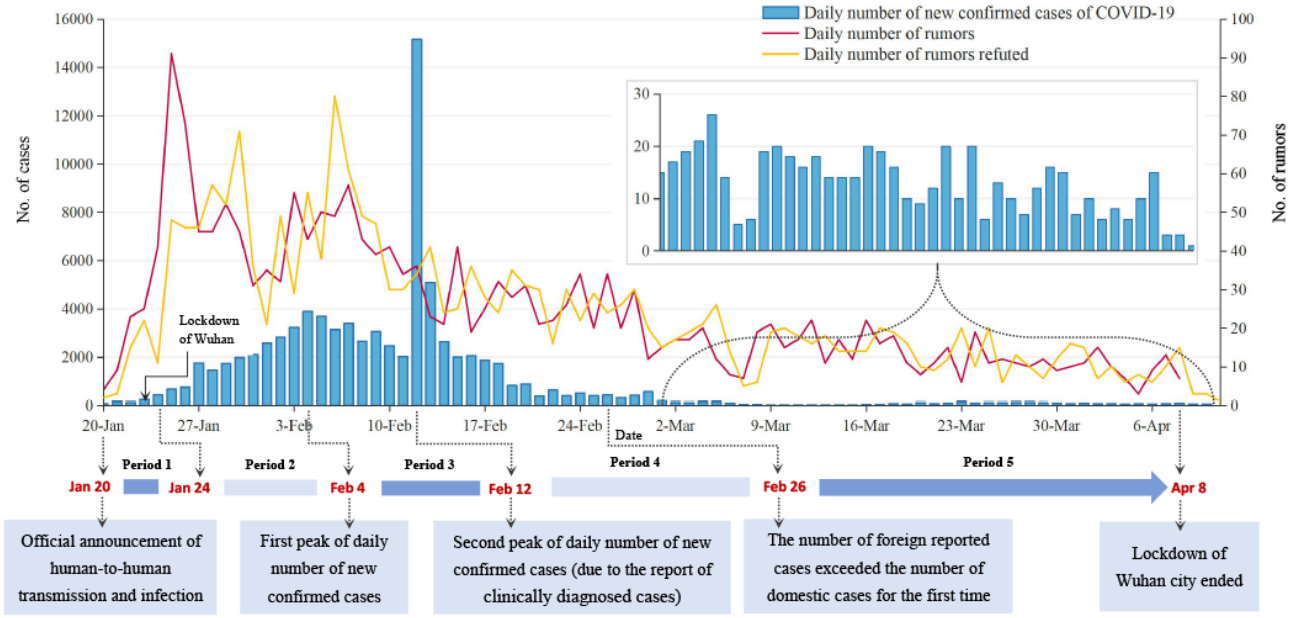
Figure 1. Flowchart of the selection of rumor posts.



Classification of the Five Time Periods

To better reflect the dynamics of the spread of rumors, five time periods were classified based on key events of the disease epidemic that could affect the dissemination of rumors on the internet (Figure 2). The time period between January 20, 2020 (the date of the first official announcement of human-to-human transmission and infection of SARS-CoV-2), and January 24, 2020, was taken as the first period in this study [20]. During this period, the National Health Commission issued a bulletin that COVID-19 had been incorporated into the management of the Prevention and Control of Infectious Diseases Act [21], and many local governments had launched first-level responses in the face of the outbreak. On January 23, the Wuhan government blocked all outbound transportation from the city with traffic suspension and home quarantine within the city [22]. The second period was January 24 to February 4; during this time, the number of new confirmed cases in China was gradually rising, and it reached its highest daily increase on February 4 [23]. From February 5 to February 12 (the third period), there was a steady decrease in the daily report of new confirmed cases. However, the secondary peak of the case numbers was reached on February 12 due to reports of clinically diagnosed cases [24]. From then until February 26 (the fourth period), the domestic epidemic gradually declined; meanwhile, foreign cases surpassed the domestic numbers for the first time by February 26. The fifth period was February 27 to April 8, when the domestic epidemic situation continued to decline and the epidemic situation in Wuhan gradually improved. On April 8, Wuhan government announced that the lockdown of the city had ended [25].

Figure 2. Daily numbers of rumors detected and refuted, the epidemic curve, and key events across five periods of the COVID-19 outbreak in China. The inset shows a magnified view of the number of COVID-19 cases over the last two periods.



Classification and Definition of Information Sources for Rumors

The rumor information sources were divided into the following categories according to their original publishing platforms: chat tools, Weibo, web pages, and others. Chat tools included WeChat and Tencent QQ. Both of these tools are major social media platforms in China that enable people to chat, communicate, and post events or opinions that others can comment on. If a rumor was first published on Weibo, the Twitter-like platform, which enables users to repost or comment, its information source was classified as Weibo. Web pages represented a number of rumors published in the form of web links. “Others” indicated some platforms that could not be classified in any of the above categories, including several mobile apps, such as TikTok, and platforms on which the source of the rumor could not be traced.

Classification and Definition of Rumor Refuters

According to the level of representativeness and influence, the rumor refuters were divided into three main categories: national, local, and other. “National” indicated that the official accounts of the rumor refuters were set up by national departments or institutions for publishing official state news or policies. Accounts established by relevant local agencies for publishing local news were classified as local, and other accounts that could not be classified in either of the above categories were classified as others; these users mainly included certified individual accounts, foreign accounts, and certified accounts of nonstate, nonlocal business firms. Meanwhile, based on the essential attributes and affiliations of different rumor refuters, all of them were further classified as government authorities, news media, and organizations, companies, or individuals. All data were classified based on the classification criteria outlined by two researchers. A third reviewer discussed the different classifications with the two reviewers and finalized the categorization. Some sample accounts are listed in Table 1.

Table 1. The criteria of the categories for the rumor refuters and the sample accounts.

Category of rumor refuter	Definition	Sample accounts
Level of representativeness and influence		
National	Official accounts set up by national departments or institutions for publishing official state news or policy	• CCTV News • Chinese National Platform to Refute Rumors • China Science Communication
Local	Accounts established by relevant local agencies for publishing local news	• Wuhan Release • Nanjing Release • Shanghai Customs
Others	Accounts that could not be classified into the categories above	• Alipay • Doctor Guan
Essential attributes and affiliations		
Government authorities	Official accounts established by government departments for the publication of government-related information	• Wuhan Release • Fuyang Health Committee • Jiangxi Public Security Department
News media	Official accounts of news media websites	• CCTV News • Wuhan Daily News • People's Daily
Organizations/companies/individuals	Accounts of some nongovernmental organizations and corporate and individual accounts.	• Beijing Public Transportation group • Ding Talk • Alipay

Analysis of Most Frequently Appearing Words in Rumors

Rumor word frequency analysis was carried out using Python version 2.7. The jieba package was employed as a text analysis tool for processing the core message of each rumor. The core message of every rumor in our study was summarized by the Rumors on Weibo account, which captured the main point of the content of the entire rumor well. Additionally, the rumors presented in the forms of images, videos, and audio were also translated by extracting the core message of that rumor through the official rumor-refuting account. Accordingly, all the rumor information in this study consisted of brief textual sentences. First, the spaces and newlines were removed from the text; then, all the punctuation marks in the text were replaced with spaces. The processed text was partitioned using the jieba package, and stop words such as "have," "is," "will," "can," and "have been" were removed from the text according to a predefined list of deactivated words. After the data set was created, a corpus of rumors was formed.

Based on the collection and classification of the rumor corpus obtained by preprocessing, the lists of words were filtered and merged according to an expert's opinion, and the data set of rumor features was finally obtained. The weight of every term was calculated using the term frequency–inverse document frequency (TF-IDF) method using the equation below:

The formula is divided into two parts; $(f_k d_k)$ to the left of the equal sign represents the word frequency, which is the number of times a word appears in the text. The higher the number of occurrences of a word, the greater the role it plays in the text. In contrast, $(f_k d_k)$ to the right of the equals sign is the logarithm value of the inverse document frequency. T represents the number of texts in the corpus, and $T(f_k)$ indicates the number of texts containing specific terms in the corpus.

The word cloud of rumors about COVID-19 during the whole study period was visualized using the wordcloud package; moreover, the summary of the top 20 high-frequency keywords of the rumors based on the term frequency–inverse document frequency (TF-IDF) values over each period was used to analyze the changes in the same keyword at different stages, which may further reflect the shifts in the public's focus.

Statistical Analysis

Excel (Microsoft Corporation) was used to record and sort information about the rumors. The epidemic curve, daily number of rumors released and clarified, new confirmed cases of COVID-19, and key events across the five periods were plotted to comprehensively analyze the relationship between the epidemic and the public's focus at different stages. Descriptive analysis of the basic information of the rumors was conducted using SPSS for Windows, version 24.0.0 (IBM Corporation). The geographic distributions of rumor sources and refuters were graphed using the pyecharts package in Python version 2.7. In addition, the cumulative number of cases and rumors in each province was calculated. The Pearson chi-square test and Fisher

exact test were performed to compare different characteristics of the rumors by each category across the five periods. The Spearman rank correlation coefficient was used to explore the relationship between the number of rumors and the epidemic trends if the variables did not satisfy the normal distribution. *P* values were 2-tailed, with statistical significance set at .05.

Results

A total of 1943 rumors were collected from the Rumors on Weibo account in this study between January 20 and April 8, 2020.

Characteristics of Rumors According to the Five Periods

The numbers of rumors published across the periods were 102, 547, 349, 377, and 568, respectively (Table 2). Rumors in the form of texts were predominant (1241/1943, 63.9%), accounting for more than half of the rumors in each time period, followed by rumors in a combination of two or more formats (330/1943, 17.0%). Among the rumor-spreading platforms, chat tools were the most common (1412/1943, 72.7%), with 1386/1943 rumors circulating in WeChat, accounting for the vast majority (98.2%). The proportions of the 1943 rumors circulating from Weibo (n=180, 9.3%), web pages (n=162, 8.3%), and other platforms (n=189, 9.7%) were similar. Additionally, 19.4% (376/1943) and 79.1% (1537/1943) of the rumors were refuted by national and local agencies, respectively. Most rumors were clarified by relevant government authorities (1250/1943, 64.3%), followed by the news media (628/1943, 32.3%).

The epidemic curve and daily number of posted and refuted rumors graphed according to the key events are illustrated in Figure 2. Spearman rank coefficient analysis showed that the daily number of posted rumors was positively associated with the daily number of new confirmed cases (Spearman rank correlation coefficient 0.73, $P<.001$). The median of the response interval between the time when the rumors were initially published and debunked was 1 day, with an IQR of 1-2. Most rumors detected were mainly concentrated between January 24 and February 7, while the majority of the refuting posts were concentrated between January 25 and February 9, with the highest daily reports of posting and refuting rumors occurring on January 25 (n=91) and February 6 (n=80), respectively.

Table 2. Characteristics of rumors across five periods during the outbreak of COVID-19 in China (N=1943).[a] $P<.001$ for all categories.

Characteristic	Time periods (2020)					Total (N=1943)	χ^2 (*df*)
	Jan 20-24 (n=102)	Jan 25-Feb 4 (n=547)	Feb 5-12 (n=349)	Feb 13- 26 (n=377)	Feb 27-Apr 8 (n=568)		
Format of rumor, n (%)							41.6 (12)
Text	69 (67.6)	354 (64.7)	228 (65.3)	228 (76.4)	362 (63.7)	1241 (63.9)	
Picture	6 (5.9)	26 (4.8)	18 (5.2)	17 (4.5)	47 (8.3)	114 (5.9)	
Multimedia[b]	6 (5.9)	85 (15.5)	38 (10.9)	41 (10.9)	88 (15.5)	258 (13.3)	
Combination[c]	21 (20.6)	82 (15.0)	65 (18.6)	91 (24.1)	71 (12.5)	330 (17.0)	
Initial platform of rumor posting, n (%)							127.6 (12)
Chat tools[d]	75 (73.5)	453 (82.8)	270 (77.4)	279 (74.0)	335 (59.0)	1412 (72.7)	
Weibo	15 (14.7)	39 (7.1)	41 (11.7)	30 (8.0)	55 (9.7)	180 (9.3)	
Web pages	10 (9.8)	29 (5.3)	11 (3.2)	29 (7.7)	83 (14.6)	162 (8.3)	
Others[e]	2 (2.0)	26 (4.8)	27 (7.7)	39 (10.3)	95 (16.7)	189 (9.7)	
Level of rumor refuter, n (%)							61.0 (8)
National	29 (28.4)	65 (11.9)	63 (18.1)	77 (20.4)	142 (25.0)	376 (19.4)	
Local	66 (64.7)	478 (87.4)	280 (80.2)	295 (78.2)	418 (73.6)	1537 (79.1)	
Others	7 (6.9)	4 (0.7)	6 (1.7)	5 (1.3)	8 (1.4)	30 (1.5)	
Type of rumor refuter, n (%)							219.5 (8)
Government authority	67 (65.7)	466 (85.2)	240 (68.8)	217 (57.6)	260 (45.8)	1250 (64.3)	
News media	25 (24.5)	74 (13.5)	101 (28.9)	152 (40.3)	276 (48.6)	628 (32.3)	
Organization, company, or individual	10 (9.8)	7 (1.3)	8 (2.3)	8 (2.1)	32 (5.6)	65 (3.3)	

[a]The five periods were classified based on key events and the disease epidemic that could affect the dissemination of rumors on the internet from January 20 to April 8, 2020.

[b]Multimedia: video, audio, and news reports.

[c]Combination: two or three formats were combined to disseminate rumors.

[d]Chat tools: WeChat and Tencent QQ.

[e]Others: other platforms that could not be classified in any of the above categories, including several mobile apps such as TikTok, and platforms that could not be traced back.

Characteristics of Rumors on Different Posting Platforms

A comparison of the 1943 rumors categorized by rumor-posting platform is shown in Table 3. Text (1241/1943, 63.9%) was the most common rumor format across different posting platforms, while the image format (114/1943, 5.9%) had the lowest percentage of all rumors. In addition to the texts, there were more rumors disseminated in a combination of formats on Weibo (60/180, 33.3%), while web pages (54/162, 33.3%) and others (42/189, 22.2%) tended to be the initial publishing platforms for rumors in multimedia format. Pearson chi-square tests indicated that there were statistically significant differences in the formats of rumors classified by platform ($\chi^2_9=142.6$, $P<.001$). Local agencies played a large role in dispelling rumors on the rumor-spreading platforms. The rumor refuters of government authorities and news media worked in tandem and complemented each other, and together they dispelled approximately 90% of the rumors on every platform. Fisher exact tests and Pearson chi-square tests suggested that the types of rumor refuters were significantly different across the platforms ($P<.001$).

Table 3. Comparison of different rumor posting platforms between January 20 and April 8, 2020 (N=1943). *P*<.001 for all categories.

Classification	Initial platform of rumor posting, n (%)					
	Chat tools[a] (n=1412)	Weibo (n=180)	Web pages (n=162)	Others[b] (n=189)	Total (N=1943)	χ^2 (*df*)
Format of rumor, n (%)						142.6 (9)
Texts	941 (66.6)	85 (47.2)	81 (50.0)	134 (70.9)	1241 (63.9)	
Pictures	91 (6.4)	11 (6.1)	8 (4.9)	4 (2.1)	114 (5.9)	
Multimedia[c]	138 (9.8)	24 (13.3)	54 (33.3)	42 (22.2)	258 (13.3)	
Combination[d]	242 (17·1)	60 (33.3)	19 (11.7)	9 (4.8)	330 (17.0)	
Level of rumor refuter, n (%)						N/A[e]
National	145 (10.3)	63 (35.0)	81 (50.0)	87 (46.0)	376 (19.4)	
Local	1256 (89·0)	109 (60.6)	73 (45.1)	99 (52.4)	1537 (79.1)	
Others	11 (0.8)	8 (4.4)	8 (4.9)	3 (1.6)	30 (1.5)	
Type of rumor refuter, n (%)						180.2 (6)
Government authorities	1030 (72.9)	87 (48.3)	56 (34.6)	77 (40.7)	1250 (64.3)	
News media	354 (25.1)	80 (44.4)	95 (58.6)	99 (52.4)	628 (32.3)	
Organization/company/individual	28 (2.0)	13 (7.2)	11 (6.8)	13 (6.9)	65 (3.3)	

[a]Chat tools: WeChat and Tencent QQ.

[b]Others: other platforms that could not be classified in any of the above categories, including several mobile apps such as TikTok, and platforms on which the source of the rumors could not be traced.

[c]Multimedia: video, audio, and news reports.

[d]Combination: two or three formats were combined to disseminate rumors.

[e]N/A: not applicable (*P* value was calculated using Fisher exact test).

Geographic Distribution of Rumor Sources and Refuters

The cumulative numbers of confirmed cases of COVID-19 and rumors generated by province across China as of April 8, 2020, are visualized in Figure 3. The internet rumors spread more actively in the southeast regions. In particular, Hubei Province and its surrounding areas showed the highest numbers of both confirmed cases and rumors in the country. The cumulative number of confirmed cases in Guangdong Province as of April 8 was ranked second in the country, whereas rumor generation in that province was relatively low. Conversely, although Guangxi Province had fewer cases, its number of rumors was among the highest nationwide.

The geographic distributions of the rumor sources and their corresponding refuters are graphed in Figure 4, where the arrow symbols indicate rumor-refuting locations. Beijing and Wuhan were the two main hubs of disinformation refuting during this stage of the COVID-19 epidemic. In Guangdong Province, where the epidemic was relatively severe, more rumors were refuted than spread. Except for the rumors that were clarified locally, rumors circulating within the provinces were often officially refuted by the provincial governments.

Figure 3. Cumulative numbers of confirmed cases of COVID-19 and rumors in each region of China as of April 8, 2020.

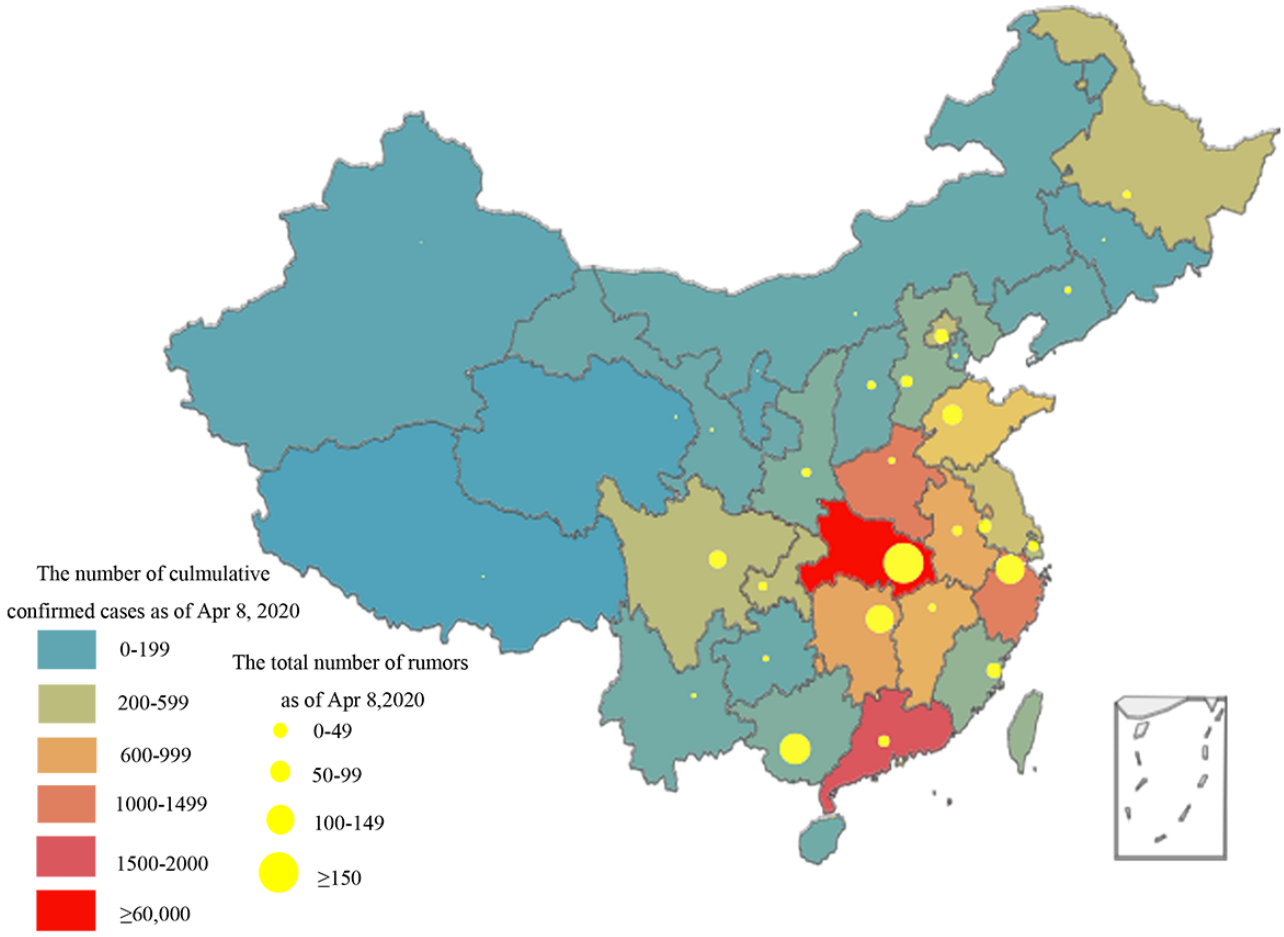


Figure 4. The geographic distributions of rumor sources and refuters during the outbreak of COVID-19 between January 20 and April 8, 2020, in China.

General Focus and Frequent Words of Rumors During the Five Phases

The top 100 most frequent words of the core messages of the rumors throughout the study period and different phases of the outbreak are depicted in Figure 5 (also see Multimedia Appendix 1 for a description of the evolution trends of the top 10 most frequent words, excluding the search keywords, at different periods during the outbreak). The average length of the textual rumors was nearly 19 characters. Overall, 冠状病毒 (coronavirus) and 肺炎 (pneumonia) were the two most common words included in the rumor core message. A majority of the words, such as 病毒 (virus), 新型 (novel), 新冠 (the Chinese abbreviation for COVID-19), and 病例 (case), were associated with COVID-19. In addition, 武汉 (Wuhan), 医院 (hospital), 小区 (residential areas), as terms related to specific locations were also frequently found in the rumors. Meanwhile, some words describing epidemic prevention and control, such as 口罩 (mask), 消毒 (disinfection), and 隔离 (quarantine), were often seen.

Figure 5. The top 100 most frequent words in the core messages of rumors during the outbreak of COVID-19 in China: (A) during the whole study period, (B) from January 20-24, 2020, (C) from January 25 to February 4, 2020, (D) from February 5-12, 2020, (E) February 13-26, 2020, and (F) from February 27 to April 8, 2020.

More details of the top 20 words listed according to their TF-IDF values in the major rumor messages over different periods are presented in Table 4. The top 5 words in the first and second periods were the same. However, the TF-IDF value of each of these words was different; the values in the first stage were higher than those in the second stage. By the third period, 网传 (sourced from internet) had become the top word, while the words 消毒 (disinfection) and 口罩 (mask) had replaced 武汉 (Wuhan) and 感染 (infection) as the fourth and fifth most common words, respectively. In contrast to similar rumors about the epidemic at the other stages, disinformation related to 开学 (schools reopen) began to grow during the fourth and fifth periods. Moreover, as the outbreak of COVID-19 increased outside China, rumors of foreign epidemic situations increased as well.

Table 4. The top 20 keywords in rumors across five periods during the outbreak of COVID-19 in China.[a]

Rank[b]	Time periods (2020)									
	Jan 20-24		Jan 25-Feb 4		Feb 5-12		Feb 13-26		Feb 27-Apr 8	
	Keyword	TF-IDF[c] value	Keyword	TF-IDF value	Keyword	TF-IDF value	Keyword	TF-IDF value	Keyword	TF-IDF value
1	冠状病毒 (coronavirus)	0.480	冠状病毒 (coronavirus)	0.256	网传 (source from internet)	0.157	新冠 (novel corona)	0.251	新冠 (novel corona)	0.283
2	肺炎 (pneumonia)	0.434	肺炎 (pneumonia)	0.237	冠状病毒 (coronavirus)	0.132	网传 (source from internet)	0.219	肺炎 (pneumonia)	0.187
3	新型 (novel)	0.294	新型 (novel)	0.182	肺炎 (pneumonia)	0.123	肺炎 (pneumonia)	0.149	网传 (source from internet)	0.163
4	武汉 (Wuhan)	0.284	武汉 (Wuhan)	0.167	消毒 (disinfection)	0.119	确诊 (confirmed diagnosis)	0.112	确诊 (confirmed diagnosis)	0.122
5	感染 (infection)	0.235	感染 (infection)	0.163	口罩 (mask)	0.085	病毒 (virus)	0.100	开学 (schools reopen)	0.117
6	网传 (source from Internet)	0.151	确诊 (confirmed diagnosis)	0.126	新型 (novel)	0.085	开学 (schools reopen)	0.099	隔离 (quarantine)	0.074
7	病例 (case)	0.127	网传 (source from internet)	0.116	确诊 (confirmed diagnosis)	0.083	隔离 (quarantine)	0.073	病例 (case)	0.070
8	确诊 (confirmed diagnosis)	0.112	疫情 (epidemic situation)	0.102	紧急通知 (urgent notice)	0.079	疫情 (epidemic situation)	0.067	病毒 (virus)	0.069
9	医院 (hospital)	0.111	隔离 (quarantine)	0.095	新冠 (novel corona)	0.076	病例 (case)	0.067	疫情 (epidemic situation)	0.065
10	新冠 (novel corona)	0.090	新冠 (novel corona)	0.077	武汉 (Wuhan)	0.075	防疫 (anti-epidemic)	0.063	人员 (personnel)	0.052
11	疑似 (suspected)	0.081	病例 (case)	0.068	隔离 (quarantine)	0.075	感染 (infection)	0.056	冠状病毒 (coronavirus)	0.046
12	患者 (patients)	0.072	一例 (one)	0.065	疫情 (epidemic situation)	0.067	冠状病毒 (coronavirus)	0.055	口罩 (mask)	0.045
13	隔离 (quarantine)	0.062	病毒 (virus)	0.062	自来水 (piped water)	0.067	口罩 (mask)	0.054	小区 (residential areas)	0.044
14	出现 (appear)	0.058	死亡 (death)	0.059	红灯 (red light)	0.063	闲逛 (hang out)	0.053	患者 (patients)	0.043
15	一例 (one)	0.058	小区 (residential areas)	0.055	酒精 (ethyl alcohol)	0.062	患者 (patients)	0.052	预防 (prevention)	0.042
16	口罩 (mask)	0.057	回来 (back)	0.052	感染 (infection)	0.057	手册 (handbook)	0.051	一例 (one)	0.041
17	疑似病例 (suspected cases)	0.047	医院 (hospital)	0.049	静置 (placed still)	0.055	小区 (residential areas)	0.051	感染 (infection)	0.038
18	预防 (prevention)	0.043	疑似 (suspected)	0.046	氯气 (chlorine)	0.054	日起 (as from today)	0.045	医院 (hospital)	0.030
19	发现 (find)	0.039	患者 (patients)	0.041	小区 (residential areas)	0.054	医院 (hospital)	0.042	武汉 (Wuhan)	0.029
20	病毒 (virus)	0.037	封城 (lockdown)	0.040	大面积 (large tracts of land)	0.053	武汉 (Wuhan)	0.041	韩国 (Korea)	0.028

[a]The five periods were classified based on key events and the disease epidemic that could affect the dissemination of rumors on the internet from January 20 to April 8, 2020.

[b]Keywords are ranked according to the TF-IDF values of the words from high to low.

[c]TD-IDF: term frequency–inverse document frequency.

Discussion

Principal Findings

Based on the official Sina Weibo rumor-refuting platform, nearly 2000 rumors that spread over the internet during the COVID-19 epidemic in China between January 20 and April 8, 2020, were investigated. This is the first domestic research to analyze the distribution, characteristics, spreading trend, and most frequent words of rumors related to the epidemic situation, which will be propitious to provide a scientific reference for the prevention and control of network rumors during unexpected events in the future.

During the study period, the median of the response interval was 1 day, indicating the timeliness of the rumor-refuting measures conducted by the Chinese government during the COVID-19 outbreak. In general, the number of rumors and refuted rumors in the first and second periods of the epidemic showed rapid growth, while both showed fluctuating declining trends in the latter three periods. Notably, after the announcement of the Wuhan lockdown, the rumor posts reached their first peak within three days, suggesting that the rapid rise in the number of rumors over the early period has a strong link with the emergence of landmark events. At this stage, due to the sudden outbreak of COVID-19, the etiology and trend of the disease were totally unclear, and the monitoring mechanism for identifying and refuting disinformation had not yet been perfected; thus, the old rumors were not quickly clarified, while new rumors appeared in rapid succession [26]. The new and old rumors intertwined to reach the peak of rumor growth, obfuscating the truth and increasing the difficulty of epidemic prevention and control [27]. By the middle of the epidemic period, most people had developed a preliminary understanding of the disease after obtaining more official information. At this stage, the rumors were less related to symbolic events and were mainly affected by the trend of the epidemic situation, fluctuating with the increase or decrease in the number of cases. Finally, by the time the disease was under control, most people had grasped a more rational understanding of the situation, with the anxiety and tension over the uncertainties greatly alleviated, leading to a decrease of the number of rumors and their corresponding clarifications. Therefore, the early stages of public emergencies, especially new infectious diseases, are the critical period for web-based surveillance of public response, risk communication, and timely release of information from credible sources, as reflected in a study on avian influenza A (H7N9) [28]. Moreover, risk communication will promote community engagement, decrease rumors to maintain social stability, and reduce threats to public health [29]. Intensive information communication with reference to hot topics of rumors may buy time to control outbreaks and reduce the risk of transmission to humans [30,31]. Additionally, transparent sharing of information in time, particularly of adverse information, and projecting uncertainty explicitly are integral parts of the management of large-scale epidemics and other emergencies [32]. Moreover, during the period of steady decline, continuous internet surveillance of rumors is still required.

Different characteristics of the rumors were analyzed in this study; it was found that internet rumors in the early stage of the epidemic were mainly disseminated in text format, most commonly in the WeChat chat tool. More processed and visualized rumors (eg, in the formats of pictures and multimedia) emerged over the later stages, while the number of rumors increased on other platforms, such as Weibo and web pages. The difference in the main formats of rumor dissemination among different platforms may be related to the openness and information screening mechanisms of each platform. In recent years, WeChat has rapidly become the main social platform in China due to its convenience and accessibility. Compared to other social media, WeChat is a social tool based on realistic relationships and closed-loop communication in a relatively private space, implying the reliability and authenticity of information and invisibly increasing the influence of rumors [33,34]. Therefore, when false news circulates, WeChat lacks self-correction ability due to the trust among acquaintances, and it is also more difficult to completely convince WeChat users that a rumor is false even with rumor-dispelling messages on the internet. However, Weibo and web pages are more open and diverse; uncertain information can be questioned, corroborated, corrected, and supplemented through user-produced content, constantly discarding false information and approaching the truth in positive interactions [35,36]. Moreover, Weibo has established an increasingly comprehensive account for refuting rumors [19]. Eventually, through the questioning of netizens, inaccurate information is replaced by the truth. In recent years, WeChat and other platforms have also taken measures to combat rumors [37]. However, the dissemination of refuting information on WeChat is restricted, mainly because it cannot reach the level of interpersonal communication, leading to small-scale transmission; thus, it can only continue to be shown “to people who do not believe rumors” [38]. Thus, a more comprehensive mechanism to encourage the dissemination of rumor-dispelling information should be developed by relevant departments in the future, accelerating the spread of credible information on WeChat and extending the influence of these departments at the same time.

Analysis of the refuters of rumors showed that during the initial outbreak of COVID-19, rumors were likely to be more nationally focused because they were relatively few in number, and the national-level rumor refuting agencies played a stronger role. When the rumors gradually started spreading locally, local authorities increased their rumor-refuting efforts; combined with the increasing enhancement of self-purification of social media, this eased the pressure on government agencies [28]. Similarly, due to the small scope and influence of rumor spread in chat tools, most were clarified by local government authorities. In contrast, rumors circulating in Weibo elicited more involvement and intervention from higher-level authorities, indicating a larger impact. These findings highlight the significance of coordinating the roles of central and local agencies in the establishment of mechanisms for refuting rumors, improving the feedback mechanisms, and maximizing the self-purification ability of social media.

According to the distribution map of the rumors, Hubei Province was the most active area for rumor breeding, which may be

related to its highly severe epidemic situation. In this study, locations with more cases were likely to generate more rumors, mainly including cities around Hubei, such as Zhejiang and Hunan. However, rumors also circulated in larger quantities in provinces with lower case numbers, such as Guangxi. Accordingly, the timely identification of rumors in regions less affected by public health emergencies is of equal importance to constant internet surveillance in areas that are more severely affected. The cross-regional rumor-refuting plot shows that Beijing was the critical center for rumor clarification. Shanghai and the provincial capitals of Guangzhou, Hunan, Sichuan, and Zhejiang also played vital roles in dispelling rumors during the epidemic, reflecting to some extent that large cities are political and media centers [39].

The high-frequency words in different periods indicated that in the pre-epidemic period, the rumors were mainly related to the disease itself, with numerous descriptions of COVID-19 contained in the core message of the rumors. In the middle of the epidemic, the rumors gradually began to be associated with prevention and control measures due to the official announcement of epidemic initiatives published by departmental agencies. When the domestic epidemic had been effectively controlled, false information about measures such as school reopening and traffic resumption spread across each region, suggesting that people gradually paid more attention to information related to policy adjustment during this period compared to the previous stages [26]. The word 网传 (source from internet) was frequently included in the titles of rumors. The word 确诊 (confirmed diagnosis) continued to appear frequently across different periods, even when the number of daily new confirmed cases had decreased since the third period. Further, when the words 确诊 (confirmed diagnosis) and 病例 (case) were combined in a rumor, the title usually took the form of “multiple cases have been confirmed in a certain place”; thus, the rumors were in a “storytelling” form that was immersive and highly convincing to readers. Such rumors have also been reported in other studies on infectious diseases [28], providing a reference for accurate identification of rumors in the future.

Limitations

There were several limitations to this study. First, this study was a retrospective analysis based on information extracted from an official rumor-refuting account, and it was difficult to avoid omitting some detailed information. In some cases, we could not trace or confirm the platforms or geographic locations where the rumors were initially published. Second, because the information of rumors in this study was refined, the emoticons and function words in the original text were not addressed; this could be considered in future research. Third, not all the rumors could be refuted on this account; instead, the most socially influential rumors were included, which could lead to information loss from the less influential rumors.

Conclusions

Our findings indicate the significance of timely management of and responses to internet rumors during major public crises. In this wave of the COVID-19 outbreak, authorities have taken effective measures to quickly dispel rumors; however, more effort could be made to better address the rumors. WeChat and other chat tools were found to be the most common origins of rumors, suggesting that the early detection and debunking mechanisms of rumors should be strengthened in closed-loop communication environments. In the early stages of the event, authorities should focus on rumors in the form of texts but should also pay more attention to other forms such as multimedia as the event progresses. The words most frequently included in the core messages of the rumors varied over different periods, which may be related to the disease itself, prevention and control measures, and social recovery; this highlights that targeted policy adjustments and timely release of official information in different phases of the outbreak should be required to prevent dissemination of internet rumors. Spread of rumors across borders needs to be controlled regardless of the intensity of the epidemic in the area. Local and national authorities should strengthen joint communication and collaboration in refuting rumors and establish a cooperative refuting mechanism based on the division of functions.

Acknowledgments

This work was supported by Nation-Zhejiang Cooperating Project (grant number WKJ-2J-1001), Zhejiang Provincial Science and Public Welfare (grant number LGF19H260004), and Zhejiang Provincial Medical and Health Project (grant number 2021KY618).

Authors' Contributions

BC, JP, and JMJ designed the study. BC, XYC, and KL analyzed the data, interpreted the results, produced the figures, and prepared the manuscript with support from JP and JMJ. BC and XYC contributed to the manuscript writing. JP and JMJ supervised the project. All authors contributed to report writing and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The evolution trends of the top 10 buzzwords (excluding the search keywords) at different periods during the outbreak.
[PNG File , 281 KB - jmir_v23i2e22427_app1.png]

References

1. Clinical management of severe acute respiratory infection when novel coronavirus (2019-nCoV) infection is suspected: interim guidance. World Health Organization. 2020 Jan 28. URL: https://apps.who.int/iris/bitstream/handle/10665/330893/WHO-nCoV-Clinical-2020.3-eng.pdf?sequence=1&isAllowed=y [accessed 2020-04-23]
2. Hui DS, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - the latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis 2020 Feb;91:264-266 [FREE Full text] [doi: 10.1016/j.ijid.2020.01.009] [Medline: 31953166]
3. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513. [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
4. Coronavirus disease (COVID-19) Situation Report – 163. World Health Organization. 2020 Jul 01. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200701-covid-19-sitrep-163.pdf?sfvrsn=c202f05b_2 [accessed 2020-07-02]
5. Barrelet C, Bourrier M, Burton-Jeangros C, Schindler M. Unresolved issues in risk communication research: the case of the H1N1 pandemic (2009-2011). Influenza Other Respir Viruses 2013 Sep 27;7 Suppl 2:114-119 [FREE Full text] [doi: 10.1111/irv.12090] [Medline: 24034495]
6. Takayasu M, Sato K, Sano Y, Yamada K, Miura W, Takayasu H. Rumor diffusion and convergence during the 3.11 earthquake: a Twitter case study. PLoS One 2015 Apr 1;10(4):e0121443 [FREE Full text] [doi: 10.1371/journal.pone.0121443] [Medline: 25831122]
7. Mocanu D, Rossi L, Zhang Q, Karsai M, Quattrociocchi W. Collective attention in the age of (mis)information. Comput Hum Behav 2015 Oct;51:1198-1204. [doi: 10.1016/j.chb.2015.01.024]
8. Global Risks 2013. World Economic Forum. URL: http://www3.weforum.org/docs/WEF_GlobalRisks_Report_2013.pdf [accessed 2020-08-04]
9. Huanqiu NW. Drinking spirit can prevent the infection of COVID-19. Webpage in Chinese. Huanqiu News Website. URL: https://china.huanqiu.com/article/9CaKrnKp2dt [accessed 2020-09-21]
10. Li J, Xie Y. Research on the influencing factors of online public opinion: an empirical analysis based on 10,600 public opinion events from 2010-2018. Article in Chinese. Journalism and Mass Communication Monthly 2020(2):37-45.
11. Zhang X, Li D, Xie G. Research on the participatory roles of media, netizens, and government in public opinion events: an empirical analysis of 3600 public opinion events. Article in Chinese. Journalism and Mass Communication Monthly 2018;6:56-63. [doi: 10.15897/j.cnki.cn51-1046/g2.2018.06.008]
12. Tchuenche JM, Dube N, Bhunu CP, Smith RJ, Bauch CT. The impact of media coverage on the transmission dynamics of human influenza. BMC Public Health 2011 Feb 25;11(S1). [doi: 10.1186/1471-2458-11-s1-s5]
13. Ma S. Analysis of rumor spreading in the microblog context of breaking events. Article in Chinese. Journalism Knowledge 2011;9:63-65.
14. Zeng R, Zhu D. A model and simulation of the emotional contagion of netizens in the process of rumor refutation. Sci Rep 2019 Oct 02;9(1):14164 [FREE Full text] [doi: 10.1038/s41598-019-50770-4] [Medline: 31578405]
15. Su Z, Li Y. Unexpected Public Health Event Calls for the Mechanism of Crisis Management in Chinese. Chinese Health Service Management 2003(9):519-520.
16. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):S154-S158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]
17. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006:244-248 [FREE Full text] [Medline: 17238340]
18. 1st WHO Infodemiology Conference. World Health Organization. 2020. URL: https://www.who.int/news-room/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference [accessed 2020-07-01]
19. Official Account to Refute Rumors. Webpage in Chinese. Sina Weibo. URL: https://weibo.com/weibopiyao?from=myfollow_all [accessed 2020-04-11]
20. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. JAMA 2020 May 19;323(19):1915-1923 [FREE Full text] [doi: 10.1001/jama.2020.6130] [Medline: 32275295]
21. Announcement No. 1. Webpage in Chinese. National Health Commission of the People's Republic of China. 2020 Jan 21. URL: http://www.gov.cn/xinwen/2020-01/21/content_5471158.htm [accessed 2020-05-13]
22. Announcement of the novel coronavirus prevention and control command in Wuhan in Chinese. National Health Commission of the People's Republic of China. URL: http://www.gov.cn/xinwen/2020-01/23/content_5471853.htm [accessed 2020-05-13]
23. National and global distribution of COVID-19 epidemic. Webpage in Chinese. Chinese Center for Disease Control and Prevention. URL: http://2019ncov.chinacdc.cn/2019-nCoV/index.html [accessed 2020-05-13]
24. Update information on COVID-19 as of February 12, 2020. Webpage in Chinese. National Health Commission of the People's Republic of China. 2020 Feb 12. URL: http://www.gov.cn/xinwen/2020-02/13/content_5478244.htm [accessed 2020-05-13]

25. Wuhan lifted its lockdown. Webpage in Chinese. Government of Hubei. 2020 Apr 08. URL: http://www.hubei.gov.cn/2019/tpyw/202004/t20200408_2207205.shtml [accessed 2020-05-13]
26. Wang C, Yu Y. The stage characteristics of rumors spread in the epidemic of COVID-19 and its countering strategies. Article in Chinese. China Newspaper Industry 2020 Apr(7):16-19.
27. Sun Y. The communication analysis of network rumor: taking the Japan Earthquake and Wenzhou Motor Car Accident as examples. Article in Chinese. Press Circles 2012(2):47-52.
28. Gu H, Chen B, Zhu H, Jiang T, Wang X, Chen L, et al. Importance of internet surveillance in public health emergency control and prevention: evidence from a digital epidemiologic study during avian influenza A H7N9 outbreaks. J Med Internet Res 2014 Jan 17;16(1):e20 [FREE Full text] [doi: 10.2196/jmir.2911] [Medline: 24440770]
29. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
30. Gu Y, Chen F, Liu T, Lv X, Shao Z, Lin H, et al. Early detection of an epidemic erythromelalgia outbreak using Baidu search data. Sci Rep 2015 Jul 28;5:12649 [FREE Full text] [doi: 10.1038/srep12649] [Medline: 26218589]
31. Chen B, Zhang JM, Jiang Z, Shao J, Jiang T, Wang Z, et al. Media and public reactions toward vaccination during the 'hepatitis B vaccine crisis' in China. Vaccine 2015 Apr 08;33(15):1780-1785. [doi: 10.1016/j.vaccine.2015.02.046] [Medline: 25731787]
32. Communicating risk in public health emergencies: a WHO guideline for emergency risk communication (ERC) policy and practice. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/259807/9789241550208eng.pdf;jsessionid=D6796FA0ECAD7D6A4E4C5456A762BDB2?sequence=2 [accessed 2020-06-18]
33. Zheng A. The Communication characteristics and governance measures of rumors in the WeChat moments. Article in Chinese. New Media Research 2015;1(12):14-15. [doi: 10.16604/j.cnki.issn2096-0360.2015.12.009]
34. Wang C. The dissemination characteristics of rumors in WeChat moments and its response atrategies. Article in Chinese. Youth Times 2017 Apr;12:97.
35. Chen B, Shao J, Liu K, Cai G, Jiang Z, Huang Y, et al. Does eating chicken feet with pickled peppers cause avian influenza? Observational case study on Chinese social media during the avian influenza A (H7N9) outbreak. JMIR Public Health Surveill 2018 Mar 29;4(1):e32 [FREE Full text] [doi: 10.2196/publichealth.8198] [Medline: 29599109]
36. Vos SC, Buckner MM. Social Media Messages in an Emerging Health Crisis: Tweeting Bird Flu. J Health Commun 2016 Dec 31;21(3):301-308. [doi: 10.1080/10810730.2015.1064495] [Medline: 26192209]
37. Under the epidemic, how to accurately dispel rumors on domestic and international social media platforms? Webpage in Chinese. The Paper. 2020 Feb 06. URL: https://m.thepaper.cn/newsDetail_forward_5776922 [accessed 2021-02-05]
38. Hu S, Dai J. The difficulty of spreading refuting information on WeChat. Article in Chinese. Journalistic Front 2017:44-45.
39. Liu K, Li L, Jiang T, Chen B, Jiang Z, Wang Z, et al. Chinese Public Attention to the Outbreak of Ebola in West Africa: Evidence from the Online Big Data Platform. Int J Environ Res Public Health 2016 Aug 04;13(8):780 [FREE Full text] [doi: 10.3390/ijerph13080780] [Medline: 27527196]

Abbreviations

H7N9: avian influenza A
SARS: severe acute respiratory syndrome
TF-IDF: term frequency–inverse document frequency

Edited by G Eysenbach; submitted 12.07.20; peer-reviewed by Y Xiang, J Lei, C Xu; comments to author 30.07.20; revised version received 18.08.20; accepted 22.01.21; published 15.02.21.

Please cite as:
Chen B, Chen X, Pan J, Liu K, Xie B, Wang W, Peng Y, Wang F, Li N, Jiang J
Dissemination and Refutation of Rumors During the COVID-19 Outbreak in China: Infodemiology Study
J Med Internet Res 2021;23(2):e22427
URL: http://www.jmir.org/2021/2/e22427/
doi:10.2196/22427
PMID:33493124

Original Paper

Telemedicine Use and Health-Related Concerns of Patients With Chronic Conditions During COVID-19: Survey of Members of Online Health Communities

Lindsey Nicole Horrell[1*], BSN, MPH, PhD; Sara Hayes[2*], MPH; Leslie Beth Herbert[2*], PhD; Katie MacTurk[2*], MS; Lauren Lawhon[2*], BS; Carmina G Valle[3,4*], MPH, PhD; Amrita Bhowmick[2,5*], MPH, MBA

[1]William F Connell School of Nursing, Boston College, Chestnut Hill, MA, United States

[2]Health Union, LLC, Philadelphia, PA, United States

[3]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[4]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[5]Department of Health Behavior, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[*]all authors contributed equally

Corresponding Author:
Lindsey Nicole Horrell, BSN, MPH, PhD
William F Connell School of Nursing
Boston College
Maloney Hall 378B
140 Commonwealth Avenue
Chestnut Hill, MA, 02467
United States
Phone: 1 617 552 4886
Email: lindsey.horrell@bc.edu

Abstract

Background: It has been widely communicated that individuals with underlying health conditions are at higher risk of severe disease due to COVID-19 than healthy peers. As social distancing measures continue during the COVID-19 pandemic, experts encourage individuals with underlying conditions to engage in telehealth appointments to maintain continuity of care while minimizing risk exposure. To date, however, little information has been provided regarding telehealth uptake among this high-risk population.

Objective: The aim of this study is to describe the telehealth use, resource needs, and information sources of individuals with chronic conditions during the COVID-19 pandemic. Secondary objectives include exploring differences in telehealth use by sociodemographic characteristics.

Methods: Data for this study were collected through an electronic survey distributed between May 12-14, 2020, to members of 26 online health communities for individuals with chronic disease. Descriptive statistics were run to explore telehealth use, support needs, and information sources, and *z* tests were run to assess differences in sociodemographic factors and information and support needs among those who did and did not use telehealth services.

Results: Among the 2210 respondents, 1073 (49%) reported engaging in telehealth in the past 4 months. Higher proportions of women engaged in telehealth than men (890/1781, 50% vs 181/424, 43%; *P*=.007), and a higher proportion of those earning household incomes of more than US $100,000 engaged in telehealth than those earning less than US $30,000 (195/370, 53% vs 241/530 45%; *P*=.003). Although 59% (133/244) of those younger than 40 years and 54% (263/486) of those aged 40-55 years used telehealth, aging populations were less likely to do so, with only 45% (677/1500) of individuals 56 years or older reporting telehealth use (*P*<.001 and *P*=.001, respectively). Patients with cystic fibrosis, lupus, and ankylosing spondylitis recorded the highest proportions of individuals using telehealth when compared to those with other diagnoses. Of the 2210 participants, 1333 (60%) participants either looked up information about the virus online or planned to in the future, and when asked what information or support would be most helpful right now, over half (1151/2210, 52%) responded “understanding how COVID-19 affects people with my health condition.”

Conclusions: Nearly half of the study sample reported participating in telehealth in the past 4 months. Future efforts to engage individuals with underlying medical conditions in telehealth should focus on outreach to men, members of lower-income households, and aging populations. These results may help inform and refine future health communications to further engage this at-risk population in telehealth as the pandemic continues.

(J Med Internet Res 2021;23(2):e23795) doi:10.2196/23795

KEYWORDS

telehealth; telemedicine; coronavirus; COVID-19; chronic disease

Introduction

According to the Centers for Disease Control and Prevention (CDC), people with underlying medical conditions (eg, autoimmune disease, chronic respiratory disease, and cancer) are at greater risk of severe illness from COVID-19 than those without pre-existing conditions [1]. Experts have released multiple guidelines to protect the health and safety of at-risk communities during this pandemic, including instructions to socially distance, stay at home as much as possible, and avoid crowds [1-3]. For communities who must continue health care appointments and treatment regimens during the pandemic, telehealth, or electronic communication and information technologies to support health care [4], has been recommended as a feasible intervention to maintain continuity of care while adhering to social distancing and stay-at-home mandates [1,5-9]. Although telehealth use rapidly increased during the initial spread of COVID-19 [10,11], little information is available regarding telehealth uptake among high-risk populations with underlying medical conditions during this pandemic.

Thus, the purpose of this study is to describe the telehealth use, resource needs, and information sources of individuals with chronic conditions during the COVID-19 pandemic. Data for this study were collected from members of online health communities hosted by the digital publishing platform, Health Union, for patients with chronic conditions (eg, communities for those with migraines, lung cancer, chronic obstructive pulmonary disease, or rheumatoid arthritis) [12]. This paper also explores differences in telehealth use among sociodemographic characteristics to identify those who may be in most critical need of support. Given the importance of understanding an audience's response in health communication evaluation [13] and the fact that increased telehealth efforts are likely to continue both during and beyond the COVID-19 pandemic [11], the information included in this paper may play a helpful role in informing future efforts to engage patients in telehealth practices.

Methods

Data Collection

Data for this study were collected in an online survey distributed through Qualtrics Survey Software (Qualtrics International Inc) to members of 26 online health communities hosted by Health Union. Each community provides patients and caregivers a digital platform to learn about their diagnosis and both provide and receive social support from peers and health care providers [12]. Between May 12-14, 2020, the survey was emailed to 98,983 community members who had previously opted into receiving emails from Health Union. The study team also posted information about the survey on the online health community websites owned and operated by Health Union. To complete the survey, respondents had to be 18 years or older, living within the United States, aware of COVID-19, and diagnosed with one of 26 chronic conditions. The survey included 35 questions assessing telehealth engagement, health information sources and needs during this pandemic, concerns about COVID-19, self-reported chronic condition diagnoses, and demographic characteristics. In this survey, telehealth was defined as "virtual appointments with doctors that may take place over video chat." Upon completion, each participant was entered into a chance drawing to win one of one hundred US $25 e-gift cards.

Measures

Telehealth Engagement and Satisfaction

In this study, telehealth engagement was measured by asking participants to respond "yes" or "no" to the question "Have you had a 'virtual visit' (telehealth) with your doctor / healthcare professional in the past 4 months?" To further assess telehealth practices, participants were asked, "You may have heard about virtual appointments with doctors (sometimes called 'telehealth') that may take place over video chat. Which statements below about telehealth apply to you?" Participants then indicated all statements that applied to them from a list of eight items such as "My doctor has reached out to me about telehealth," "I am interested in telehealth but don't know how to use it," and "I started using telehealth (or will start) because of the pandemic." The full list of item responses can be found in Multimedia Appendix 1.

Among those who had participated in a telehealth visit in the past 4 months, satisfaction was assessed by asking participants to rate their agreement with the following three statements: "I had a positive experience using telehealth," "The technology was difficult to use," and "I feel like the virtual visit was just as good (or better) than an in-person visit." Responses were on a 7-point Likert scale (1=completely disagree to 7=completely agree); in these analyses, "agreement" with each statement was defined as selecting a 6 or 7, while "disagreement" was defined as selecting a 1 or 2.

COVID-19 Information Sources and Needs

COVID-19 Communications With a Health Care Team

To assess how individuals with underlying medical conditions have been gathering information about COVID-19, participants were first asked questions about pandemic-related interactions with their health care team. For example, participants were

asked, "In which of the following ways have you communicated with your doctor/healthcare team about the novel coronavirus (COVID-19)?" and prompted to choose from a list of eight item responses (eg, I reached out, my doctor personally contacted me, or we discussed it at a regular appointment). Among those who had communicated with their health care team, participants were asked, "How have you communicated with your doctor/healthcare professional?" (eg, by phone, portal, or email) and "What did you discuss with your doctor/healthcare provider?" (eg, deciding to temporarily stop or skip medications, monitoring for COVID-19 symptoms, or regular discussion or check-in about my current condition and symptoms). Participants were also asked "What, if anything, have you done – or are you planning to do in the near future -- in response to hearing about the novel coronavirus (COVID-19)?" Item responses included "cancel or postpone regular doctor/healthcare visits"; "have a 'virtual' visit with a doctor/healthcare professional – 'telehealth' – via video chat"; and other items related to health decisions, needs, and information seeking behavior. A complete list of item responses for each of these questions can be found in Multimedia Appendix 1.

Patient Concerns During the COVID-19 Pandemic

To assess patient concerns during the COVID-19 pandemic, participants were asked to indicate their agreement with a series of nine statements on a 7-point Likert scale (1=completely disagree to 7=completely agree). The full list of statements can be found in Multimedia Appendix 1, but examples included "Having a chronic health condition makes me feel particularly concerned about the coronavirus," "I feel like people are not taking the coronavirus seriously enough," and "I feel like I am taking all the right precautions to reduce my risk of getting coronavirus." Participants who reported currently or previously having cancer were asked to rate an additional statement: "Having (or having had) cancer makes me feel particularly concerned about the coronavirus." Agreement with each statement was again defined as selecting a 6 or 7, while disagreement was defined as selecting a 1 or 2.

Other COVID-19–Related Information Sources and Needs

To identify other sources this population has turned to for information during the pandemic, participants were asked, "What sources are you using to learn more about the novel coronavirus (COVID-19)?" and prompted to select all that apply from a list of 15 items (eg, internet search engines, social networking sites, or TV news reports). Finally, participants were asked to indicate the information or support that would be most helpful to them right now by selecting up to three choices from a list of 10 items such as "Information/guidance from my doctor about COVID-19 and my health condition/its treatment," "emotional and/or mental health support," and "Financial support for other necessities/bills (e.g., food, rent, etc.)." The full list of item responses can be found in Multimedia Appendix 1.

Demographic and Disease Characteristics

Participants were asked to indicate their age from a dropdown menu ranging from "Under 18" to "90 or older," with each age in between listed continuously (eg, 18, 19, 20, ...). Participants were also asked to indicate their residence type (ie, rural, urban, or suburban), gender, highest level of education attainment, annual household income, and primary health insurance from categorical lists of item responses (see Multimedia Appendix 1). Participants indicated their chronic condition by selecting all that apply from a list of 25 chronic conditions, or participants were able to respond "none of these." The full list of chronic conditions can be found in Multimedia Appendix 1, but examples include asthma, cystic fibrosis, Chron disease, migraines, and hypertension. Participants were also asked to indicate if they had been diagnosed with any from a list of nine types of cancer or "other" if they had been diagnosed with a cancer that was not listed.

Analyses

Descriptive analyses were conducted to assess telehealth use, resource needs, and information sources among all participants, and *z* tests were used to explore differences between demographic groups. To assess for meaningful differences between demographic groups, age and education were each categorized into three groups (<40 years, 40-55 years, and ≥56 years, and high school, GED [General Educational Development], or less than high school; college degree; trade or vocational training; and some college, graduate, or professional degree), and income was categorized into four groups (<US $30,000, US $30,000-US $54,999, US $55,000-US $99,999, and ≥US $100,000). *Z* tests were also used to explore telehealth use by diagnosis and patient concerns.

Results

Participants

Among those who received information about the survey via email (n=98,983), 1758 (1.8%) people completed it, and an additional 452 people participated after finding information about the survey on one of the health community websites. Of the 2210 total survey respondents, 81% (n=1781) were female, 68% (n=1500) were older than 55 years, and 87% (n=1920) completed at least some college or trade school. The most commonly reported chronic conditions were hypertension (n=863, 39%), hyperlipidemia (n=562, 25%), asthma (n=424, 19%), and migraine (n=420, 19%), and 98% (n=2164) of respondents reported having health insurance. Approximately 30% (n=660) of respondents reported either full- or part-time employment, and 3% (n=75) were currently seeking employment. The remaining respondents were retired (n=849, 38%), stay-at-home parents (n=71, 4%), on disability (n=456, 21%), unemployed and not looking for employment (n=81, 4%), or currently in school (n=18, 1%). Among respondents who reported household incomes (n=1877), the majority (n=1507, 80%) reported annual household incomes under US $100,000. Table 1 displays full demographic and disease characteristics of the study sample.

Table 1. Sample demographic and chronic condition diagnoses.

Demographic	Participants (N=2210), n (%)
Age (years)	
<40	224 (10.14)
40-55	486 (21.99)
≥56	1500 (67.87)
Gender	
Female	1781 (80.59)
Male	424 (19.19)
Nonbinary/gender nonconforming	5 (0.23)
Household income (US $; n=1877)	
<30,000	530 (28.24)
30,000-54,999	428 (22.80)
55,000-99,999	549 (29.25)
≥100,000	370 (19.71)
Education	
High School/GED[a] or less than high school	290 (13.12)
College degree, trade/vocational training, or some college	1454 (65.79)
Graduate or professional degree	466 (21.09)
Employment status	
Retired	849 (38.42)
Employed	660 (29.86)
On disability	456 (20.63)
Not employed outside of the home	245 (11.09)
Health insurance	
Medicare	1040 (47.06)
Private, Employer-provided, or Health Insurance Exchange	964 (43.62)
Medicaid	130 (5.88)
Other/not sure	30 (1.36)
Do not have	46 (2.08)
Chronic condition	
High blood pressure	863 (39.05)
High cholesterol/hyperlipidemia	562 (25.43)
Asthma	424 (19.19)
Migraine	420 (19.00)
Multiple sclerosis	300 (13.57)
Rheumatoid arthritis	315 (14. 25)
COPD[b]/emphysema/chronic bronchitis	394 (17.83)
Irritable bowel syndrome	333 (15.07)
Type 2 diabetes	252 (11.40)
Psoriatic arthritis	169 (7.65)
Atopic dermatitis/eczema	111 (5.02)
Endometriosis	155 (7.01)

Demographic	Participants (N=2210), n (%)
Crohn disease	147 (6.65)
Parkinson disease	130 (5.88)
Plaque psoriasis	122 (5.52)
Lupus	82 (3.71)
Macular degeneration	125 (5.66)
Ankylosing spondylitis	91 (4.12)
Heart failure	122 (5.52)
Ulcerative colitis	98 (4.43)
Hepatitis C	51 (2.31)
Axial spondyloarthritis/nonradiographic axial spondylarthritis	22 (1.0)
Cystic fibrosis	47 (2.13)
HIV	20 (0.90)
Alzheimer disease	4 (0.18)
Cancer	791 (35.79)
None	115 (5.20)

[a]GED: General Educational Development.

[b]COPD: chronic obstructive pulmonary disease.

Telehealth Engagement and Satisfaction

Among the 2210 survey respondents, 1073 (49%) respondents reported participating in a virtual visit (telehealth) with a doctor or health care provider (HCP) in the past 4 months, while 997 (45%) canceled or postponed regularly scheduled visits with their HCP, and 809 (37%) cancelled or postponed routine medical tests or planned to do so in the near future. Of those who used telehealth services (n=1073), 68% (n=725) agreed with the statement "I had a positive experience using telehealth." Furthermore, 39% (n=420) agreed that "the virtual visit was just as good (or better) than an in-person visit," but 14% (n=145) agreed with the statement "the technology was difficult to use."

As displayed in Figure 1, 40% (n=886) of the 2210 participants started using or planned to use telehealth options because of the COVID-19 pandemic, and 9% (n=194) reported engaging in telehealth options before the pandemic. Almost half of the participants (n=945, 43%) wanted to return to face-to-face appointments in the future, but 27% (n=591) reported a desire to use telehealth even after the pandemic subsides. Finally, 7% (n=162) of participants reported an interest in telehealth but did not know how to use it, while 4% (n=82) had never heard the term "telehealth," and 7% (n=162) reported no interest in using telehealth.

Looking at telehealth use by disease characteristics and patient concerns, patients with cystic fibrosis (n=33/47, 70%), lupus (n=55/82, 67%), and ankylosing spondylitis (n=60/91, 66%) recorded the highest proportions of patients using telehealth in the past 4 months when compared to patients with other diagnoses. When compared to those who did not engage in telehealth, higher proportions of those who used telehealth services reported feeling at higher risk of severe disease due to COVID-19 as indicated by agreeing with the statements "I feel like I am at a greater risk of getting coronavirus because of the medications I take" (341/1073, 32% vs 288/1137, 25%; *P*=.001) and "I feel like I am at a greater risk of having a more severe case of coronavirus because of my general health" (676/1073, 63% vs 659/1137, 58%; *P*=.02). Higher proportions of those who engaged in telehealth also agreed with the statements "Having a chronic health condition makes me feel particularly concerned about the coronavirus" (721/1024, 70% vs 698/1071, 65%; *P*=.01), "I am concerned that I may not have access to my medication/treatment in the future" (194/1073, 18% vs 153/1137, 13%; *P*=.003), and "I feel like people are not taking the coronavirus seriously enough" (646/1073, 60% vs 630/1137, 55%; *P*=.02). Figure 2 displays percentages of those who indicated various concerns by telehealth use status.

Figure 1. Participant's self-reported experience using telehealth (N=2210).

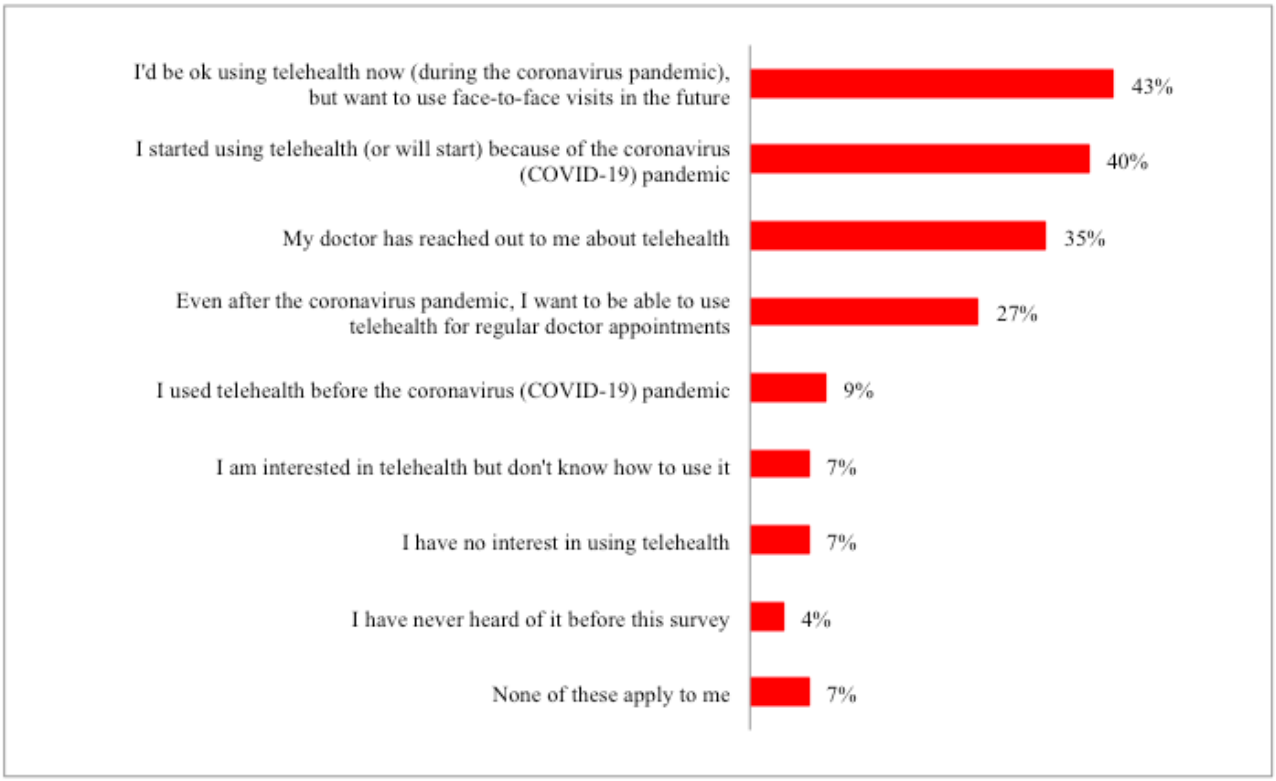
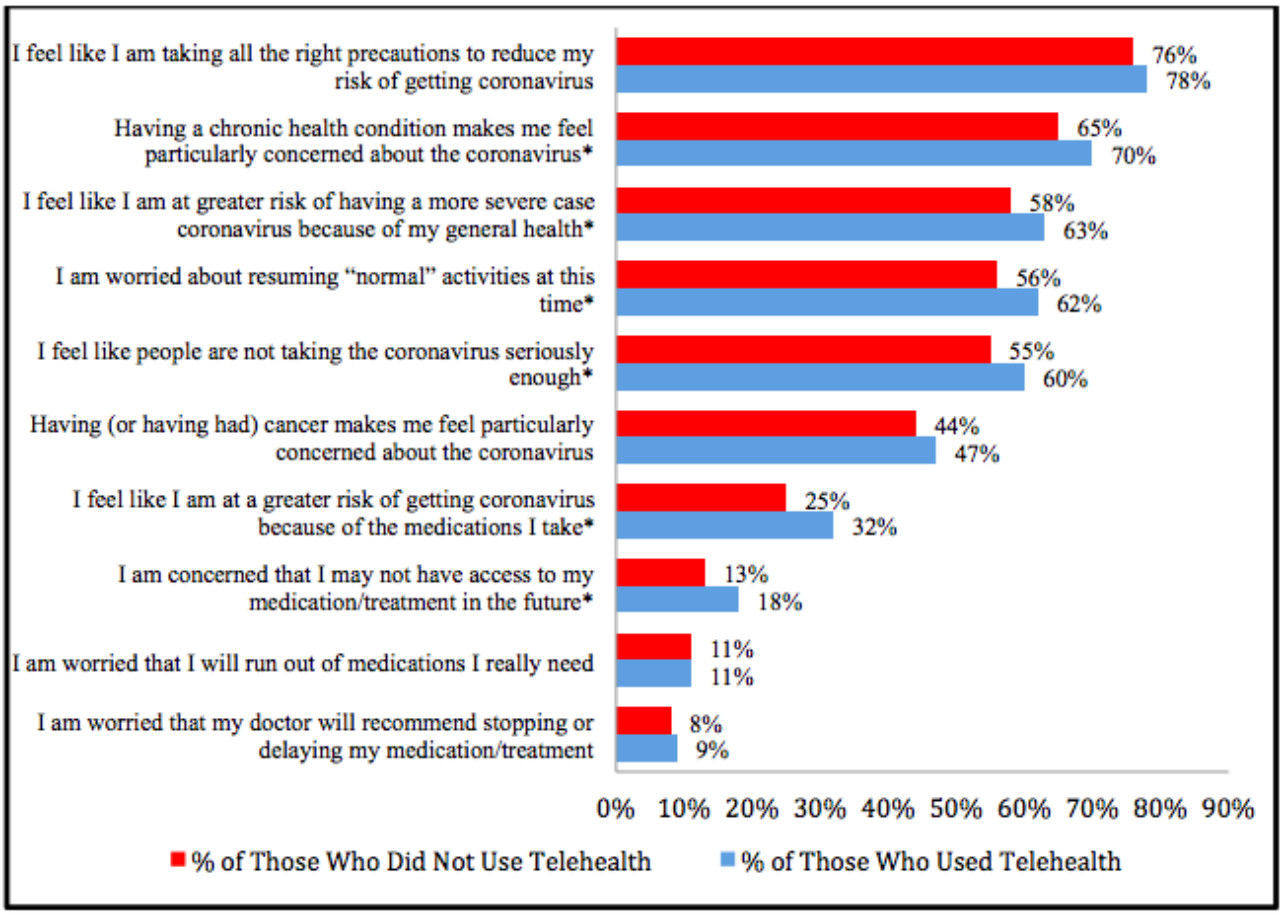


Figure 2. Participant concerns among those who used telehealth and did not use telehealth. *P<.05.



COVID-19 Information Sources and Needs

COVID-19 Communications With a Health Care Team

Of the 2210 participants, 72% (n=1592) reported communicating with their health care team about COVID-19 in some manner. A complete list of the ways in which this discussion was initiated is listed in Multimedia Appendix 2; however, the most highly endorsed responses were "I reached out via phone, online portal, etc." (n=800, 36%) and "My doctor's office sent out information and recommendations (via email, online portal, letter, social media" (n=602, 27%). The most common modes of communication among those who had spoken with an HCP about the virus were by phone (n=1183, 74%), in a telehealth appointment (n=705, 44%), and via online portal (n=680, 43%). Primary topics of these communications were decisions regarding treatment regimens (continuing as prescribed: n=933, 60%; changing: n=181, 12%; or temporarily stopping: n=119, 8%), regular check-ups (n=853, 55%), CDC COVID-19 guidelines (n=660, 42%), and decisions to cancel or delay upcoming appointments and surgeries (n=445, 29%). A complete list of topics discussed can be found in Multimedia Appendix 2.

Other COVID-19–Related Information Sources and Needs

HCPs were not the only source of information respondents turned to in response to hearing about COVID-19. When asked what sources respondents used to learn more about the virus, 63% (n=1389) selected TV news reports, 58% (n=1284) chose news websites, 57% (n=1268) turned to government websites, and 40% (n=892) reported retrieving information through search engines such as Yahoo or Google. In total, 60% (n=1333) of respondents reported that they either already looked up information online or planned to in response to hearing about COVID-19.

When asked what information or support would be most helpful right now, just over half of the survey respondents (n=1151, 52%) answered "understanding how COVID-19 affects people with my health condition," 48% (n=1071) indicated obtaining up-to-date and accurate information about the virus, 24% (n=538) selected emotional or mental health support, and 22% (n=485) indicated that receiving information from their HCP about COVID-19 and their health condition would be helpful. Additionally, 19% (n=412) indicated a need for financial support for life necessities such as food and rent. Significantly higher proportions of those who engaged in telehealth indicated a need for the following types of support than those who did not engage in telehealth: information from the company who makes my medications about COVID-19 and the treatment I take (126/1073, 12% vs 94/1137, 8%; *P*=.006), emotional or mental health support (292/1073, 27% vs 246/1137, 22%; *P*=.002), and financial support for medications and health care costs (167/1073, 16% vs 132/1137, 12%; *P*=.007). Finally, 16% (n=359) of the 2210 respondents indicated no need for additional information or support.

Demographic Comparisons

In this section, we highlight notable trends and differences in telehealth engagement, satisfaction, information sources, and needs between members of various demographic backgrounds. A full list of comparisons can be found in Multimedia Appendix 2.

Telehealth Engagement and Satisfaction

When looking at telehealth engagement across sociodemographic categories, a higher proportion of women engaged in telehealth than men (890/1781, 50% vs 181/424, 43%; *P*=.007), and a higher proportion of those earning household incomes over US $100,000 engaged in telehealth more than those earning less than US $30,000 (195/370, 53% vs 241/530, 45%; *P*=.03). The proportion of those engaging in telehealth also increased across educational attainment levels, such that a greater percentage of those who completed at least some college or trade school engaged in telehealth than those who completed high school or below (706/1454, 49% vs 114/290 39%; *P*=.004), and a greater percentage of those who completed an advanced degree engaged in telehealth than those who completed college or trade school (253/466, 54% vs 706/1454, 49%; *P*=.03). Although 59% (133/224) of those younger than 40 years and 54% (263/486) of those aged 40-55 years used telehealth, a smaller proportion of aging populations did so, with only 45% (677/1500) of individuals 56 years or older reporting telehealth use (*P*<.001 and *P*=.001, respectively). Telehealth use did not significantly differ across residence types (rural, suburban, or urban).

Higher proportions of female telehealth users found the virtual visit to be just as good or better than an in-person visit compared to males (360/890, 40% vs 58/181, 32%; *P*=.04), and a higher proportion of those with a high school degree or below also found the virtual visit to be just as good or better than an in-person visit compared to those with advanced degrees (54/114, 47% vs 91/253, 36%; *P*=.04). When comparing telehealth satisfaction among various income groups, higher percentages of those with lower household incomes (<US $30,000: 100/241, 41%; US $30,000-US $54,999: 86/208, 41%; and US $55,000-US $99,999: 108/266, 41%) also found the virtual visit to be just as good or better than an in-person visit when compared to those earning over US $100,000 annually (59/195, 30%; all *P*=.02). A greater percentage of individuals with advanced degrees disagreed with the statement "the technology was difficult to use" (184/253, 73%) compared to those who reported lower educational levels (at least some college or trade school: 441/706, 62%; *P*=.003; high school or below: 69/114, 61%; *P*=.02). Finally, a greater percentage of those in suburban areas also disagreed with this statement compared to those living in urban areas (372/549, 68% vs 135/224, 60%; *P*=.047).

A significantly larger proportion of those who were 56 years or older reported no interest in using telehealth (128/1500, 9%) compared to younger age groups (25/486, 5% of those aged 40-55 years; *P*=.02; 9/224, 4% of those younger than 40 years; *P*=.02) and desired to return to in-person appointments after the pandemic (675/1500, 45% vs 192/486, 40% of those aged 40-55 years; *P*=.03; and 78/224, 35% of those younger than 40 years; *P*=004). Higher proportions of those 56 years or older also rated the technology difficult to use (107/677, 16%, compared to 12/133, 9% of those younger than 40 years; *P*=.04; and 26/263, 10% of those aged 40-55 years; *P*=.02), but there were no

significant differences between age groups in their agreement with the statement "I feel like the virtual visit was just as good (or better) than an in-person visit." Furthermore, a larger percentage of the older age group were also interested in telehealth but did not know how to use it (125/1500, 8% vs 27/486, 6% of those aged 40-55 years; *P*=.045, and 10/224, 4% of those younger than 40 years; *P*=.04).

COVID-19 Information Sources and Needs

COVID-19 Communications With a Health Care Team

When looking at engagement with a health care team regarding COVID-19 information sources and needs, it may be notable that smaller proportions of those who earned a high school degree or less reported that their doctor's office sent out information and recommendations about the virus when compared to peers who reported higher educational attainment (53/290, 18% vs 397/1454, 27% of those who completed at least some college or trade school; *P*=.001; and 152/466, 33% who held an advanced degree; *P*<.001). Similarly, smaller proportions of individuals with household incomes under US $30,000 (134/530, 25%) and between US $30,000 and US $54,999 (109/428, 25%) reported such a distribution of information compared to those earning at least US $100,000 (122/370, 33%; *P*=.01 and *P*=.02, respectively). Furthermore, higher proportions of those younger than 40 years (113/224, 50%) and between 40 and 55 years (211/486, 43%) reached out to their HCP than those older than 55 years (476/1500, 32%; both *P*<.001). Even when meeting with an HCP, a smaller percentage of those older than 56 years discussed monitoring for COVID-19 symptoms (140/1039, 13% vs 63/350, 18% of those 40-55 years; *P*=.04; and 36/171, 21% of those younger than 40 years; *P*=.009) and what to do if experiencing COVID-19 symptoms (199/1039, 19% vs 106/350, 30% of those aged 40-55 years; *P*<.001; and 52/171, 30% of those younger than 40 years; *P*=.001) compared to younger age groups (see Multimedia Appendix 2).

Other COVID-19–Related Information Sources and Needs

Looking at which COVID-19 information sources were most popular among various demographic groups, government websites (eg, CDC, National Institutes of Health, and local health departments) were the most popular selection among participants younger than 40 years (n=165, 74%), aged 40-55 years (n=329, 68%), and those with annual household incomes over US $100,000 (n=240, 65%). TV news reports were most popular among those 56 years or older (n=1009, 67%), all those earning under $100,000 annually, and all those reporting education levels up to a college degree. They were also the most popular choice regardless of residence type and gender (see Multimedia Appendix 2). Finally, news websites were the most popular choice among those with an advance degree (n=322, 69%).

"Understanding how COVID-19 affects people with my health condition" was of primary interest across all age groups, residence types, education levels, and household incomes, with the exception of those with a high school degree or below, among whom a slightly higher percentage (135/290, 47% vs 134/290, 46%) indicated a general need for up-to-date and accurate information about COVID-19. There were significant differences in information and support needs between age groups such that higher percentages of those younger than 40 years expressed interest in the following resources compared to older age groups: financial support for necessities (69/224, 31% vs 215/1500, 14% of those 56 years or older; *P*<.001); learning how to have food, medication, or supplies delivered (29/224, 13% vs 34/486, 7% of those aged 40-55 years; *P*=.01; and 112/1500, 7% of those 56 years or older; *P*=.005); and emotional or health support (84/224, 38% vs 138/486, 28% of those aged 40-55 years; *P*=.02; and 316/1500, 21% of those 56 years or older; *P*<.001). Conversely, a higher percentage of those 56 years or older stated that they were not in need of any more information or support (280/1500, 19%) compared to those aged 40-55 years (63/486, 13%; *P*=.004) and younger than 40 years (16/224, 7%; *P*<.001).

When examining information and resource needs across other demographic factors, a significantly larger proportion of females also expressed a need for emotional or mental health support than males (465/1781, 26% vs 71/424, 17%; *P*<.001). A smaller proportion of those with an advanced degree expressed a need for financial support compared to those of lower educational attainment, both for medications and health care expenses (44/466, 9% vs 214/1454, 15% of those with some college or trade school; *P*=.004; and 41/290, 14% of those with a high school degree or below; *P*=.047), and for other necessities and bills (59/466, 13% vs 283/1454, 19% of those with some college or trade school; *P*=.001; and 70/290, 24% of those with a high school degree or below; *P*<.001). Finally, the need for information regarding COVID-19 in general and its relation to one's health condition and current treatments tended to increase as household income increased, while financial and food medication delivery information needs tended to decrease (see Multimedia Appendix 2).

Discussion

Principal Results

The purpose of this study is to assess the telehealth use, resource needs, and information sources of individuals living with chronic conditions during the COVID-19 pandemic. According to these results, 45% (n=997/2210) of the study sample either cancelled or planned to cancel appointments with their HCP during times of COVID-19, and nearly half (1073/2210, 49%) engaged in telehealth appointments in the 4 months prior to data collection. Although 40% (886/2210) of participants initiated telehealth care because of the pandemic, 7% (162/2210) responded that they were interested in telehealth and did not know how to use it, 7% (162/2210) reported that they had no interest in using telehealth, 4% (82/2210) reported they had never heard the term prior to taking the survey, and only 35% (780/2210) of patients reported that their doctor reached out to them about telehealth. Thus, as recommendations continue for patients with underlying medical conditions to mitigate possible exposure outside of the home and as practices look to extend telehealth services beyond the pandemic [11], providers need to ensure adequate information is distributed to inform patients about telehealth procedures. A recent survey of 2949 adults living in China found that high-quality communication with a provider prior to

COVID-19 was associated with preventive behavior uptake during the pandemic [14]. Distributing patient-centered, personalized communications about telehealth from HCPs may help increase uptake of telehealth and maintain continuity of care in the future.

It may also be important to note that just under half of survey respondents (945/2210, 43%) were content using telehealth services during the pandemic but desired to return to in-person care in the future. With this in mind, information provided to patients about telehealth may also include information about future plans to return to in-person visits, including public and personal health standards that need to be met to do so. In the interim, providers may also consider implementing recently suggested strategies to enhance feelings of "copresence" or connection between patient and provider during telehealth visits [7,15]. Such strategies include postvisit mood evaluations, one-click responses from health care teams, and encouraging messages, and may help further engage patients, alleviate stress, and maintain continuity of care during the pandemic [7,15].

As telehealth outreach increases, HCPs can optimize virtual visits by addressing patients' top concerns during the pandemic. When asked what support or information would be most helpful, the top-rated responses included information about COVID-19, how it impacts individuals with specific diagnoses and disease regimens, and mental or emotional health support. Experts have already noted the important role telehealth could play in providing remote interdisciplinary care [16], including mental and emotional health services [17] during this global crisis. Thus, telehealth services across multiple disciplines and hospital services, including mental health care, social work, and patient navigation, may prove vital to providing quality care and meeting all information and supportive care needs of patients during COVID-19.

Finally, it is imperative to note which groups may be at highest risk of losing access to care as services shift to digital formats during COVID-19. According to these results, greater proportions of adults younger than 56 years have engaged in and are interested in telehealth than those 56 years or older, which may be unsurprising given the reported digital divide between younger and older generations [18]. This phenomenon has also been reflected in other literature assessing telehealth use both prior to and during the pandemic [19-21]. Although a higher percentage of older adults expressed no interest in telehealth compared to younger age groups, nearly half of those 56 years or older started using telehealth in response to COVID-19, nearly half said they would be willing to participate in telehealth even though they wanted to return to in person in the future, one-third wanted to participate in telehealth even after the pandemic, and only 5% reported being interested but not knowing how to use telehealth. Given the fact that older adults, with or without underlying medical conditions, are at higher risk of serious illness from COVID-19 [1,22-24], particular efforts to communicate the benefits and importance of telehealth during the pandemic and engage this group in telehealth efforts may be vital. Because lower percentages of aging adults reported reaching out to their health care provider about COVID-19, it may also be particularly important for health care providers to initiate conversations about telehealth and pandemic-related information.

Similarly, smaller percentages of participants who reported lower household incomes and educational attainment participated in telehealth when compared to peers of higher income and educational status, realities that have also been noted in previous literature on digital health use [18,20,25]. Other researchers have noted measures HCPs can take to try to ensure telehealth does not exacerbate health inequities during the pandemic (eg, ensuring language interpreter access, recruiting telehealth staff from diverse backgrounds, providing patient trainings in telehealth use, and informing patients about free or discounted broadband access) [25,26]. Thus, future outreach efforts should focus on equitable communication and engagement of patients in telehealth practices.

Limitations

This study presents important information regarding telehealth use, resource needs, and information sources of individuals considered at risk for severe disease during COVID-19 due to underlying medical conditions. This data collection is not without limitations. Given the exploratory nature of this study, many of these measures have not been validated, and the general description of telehealth as a "virtual visit with a doctor/healthcare professional" may not encompass the full scope of telehealth practices taking place during the COVID-19 pandemic. Despite this fact, only 4% of individuals in the study reported that they had never heard of telehealth prior to the study, indicating a general awareness if the terminology among the study population. Because all participants were members of online health communities, the telehealth and information-seeking practices of this sample may also be skewed, as participants may be more comfortable with digital health solutions than the general public.

Finally, at approximately 2%, the response rate for this survey was low, the sample was primarily female (1781/2210, 81%), and the data was skewed toward older populations who rated their conditions with higher severity, which may skew the age analyses presented and limit the generalizability of these results. With this said, the demographic characteristics of this sample do mirror those of the larger sampling frame, given that, among 36,515 survey responses collected from American Health Union members over the previous year, 81% of responses were female and 48% were 60 years or older. By leveraging existing online health communities for data collection, however, this study included a large, national sample of individuals with underlying medical conditions and provides preliminary guidance regarding the telehealth practices and needs of this high-risk population during the COVID-19 pandemic.

Conclusions

These results offer important insight to current telehealth practices and information needs of patients with underlying medical conditions during COVID-19. Although many patients are engaging in telehealth, they continue to seek information about COVID-19, how the virus impacts people with their particular condition and treatment regimen, and support for mental and emotional health during the pandemic. Moving

forward, medical and public health professionals may continue to take an active approach to engaging patients with underlying conditions in available telehealth services, particularly those who are members of lower socioeconomic status and aging populations. Future research will explore changes in telehealth practices and information needs over time by examining data from multiple waves of this survey distributed between March and July of 2020.

Acknowledgments

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number T32CA128582.

Conflicts of Interest

SH, LBH, KM, LL, and AB are employees of Health Union, LLC.

Multimedia Appendix 1
Measures.
[DOCX File , 42 KB - jmir_v23i2e23795_app1.docx]

Multimedia Appendix 2
Demographic comparisons.
[XLSX File (Microsoft Excel File), 122 KB - jmir_v23i2e23795_app2.xlsx]

References

1. National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases. People with certain medical conditions. Centers for Disease Control and Prevention. 2020 Aug 14. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/groups-at-higher-risk.html [accessed 2020-08-23]
2. Güner R, Hasanoğlu I, Aktaş F. COVID-19: prevention and control measures in community. Turk J Med Sci 2020 Apr 21;50(SI-1):571-577. [doi: 10.3906/sag-2004-146] [Medline: 32293835]
3. Coronavirus disease (COVID-19) advice for the public. World Health Organization. 2020 Jun 04. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [accessed 2020-08-23]
4. The American Telemedicine Association. Telemedicine, telehealth and health information technology. World Health Organization. 2006. URL: https://www.who.int/goe/policies/countries/usa_support_tele.pdf?ua=1 [accessed 2020-10-16]
5. Burki TK. Cancer guidelines during the COVID-19 pandemic. Lancet Oncol 2020 May;21(5):629-630 [FREE Full text] [doi: 10.1016/S1470-2045(20)30217-5] [Medline: 32247319]
6. Hakim AA, Kellish AS, Atabek U, Spitz FR, Hong YK. Implications for the use of telehealth in surgical patients during the COVID-19 pandemic. Am J Surg 2020 Jul;220(1):48-49 [FREE Full text] [doi: 10.1016/j.amjsurg.2020.04.026] [Medline: 32336519]
7. Liu N, Huang R, Baldacchino T, Sud A, Sud K, Khadra M, et al. Telehealth for noncritical patients with chronic diseases during the COVID-19 pandemic. J Med Internet Res 2020 Aug 07;22(8):e19493 [FREE Full text] [doi: 10.2196/19493] [Medline: 32721925]
8. Singh AK, Gupta R, Ghosh A, Misra A. Diabetes in COVID-19: prevalence, pathophysiology, prognosis and practical considerations. Diabetes Metab Syndr 2020;14(4):303-310 [FREE Full text] [doi: 10.1016/j.dsx.2020.04.004] [Medline: 32298981]
9. Al-Shamsi HO, Alhazzani W, Alhuraiji A, Coomes EA, Chemaly RF, Almuhanna M, et al. A practical approach to the management of cancer patients during the novel coronavirus disease 2019 (COVID-19) pandemic: an international collaborative group. Oncologist 2020 Jun;25(6):e936-e945. [doi: 10.1634/theoncologist.2020-0213] [Medline: 32243668]
10. Anthony Jnr B. Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. J Med Syst 2020 Jun 15;44(7):132 [FREE Full text] [doi: 10.1007/s10916-020-01596-5] [Medline: 32542571]
11. Contreras CM, Metzger GA, Beane JD, Dedhia PH, Ejaz A, Pawlik TM. Telemedicine: patient-provider clinical engagement during the COVID-19 pandemic and beyond. J Gastrointest Surg 2020 Jul;24(7):1692-1697 [FREE Full text] [doi: 10.1007/s11605-020-04623-5] [Medline: 32385614]
12. Green BM, Van Horn KT, Gupte K, Evans M, Hayes S, Bhowmick A. Assessment of adaptive engagement and support model for people with chronic health conditions in online health communities: combined content analysis. J Med Internet Res 2020 Jul 07;22(7):e17338 [FREE Full text] [doi: 10.2196/17338] [Medline: 32492651]
13. Making health communication programs work. National Cancer Institute. 2004. URL: http://www.cancer.gov/publications/health-communication/pink-book.pdf [accessed 2020-08-23]

14. Niu Z, Wang T, Hu P, Mei J, Tang Z. Chinese public's engagement in preventive and intervening health behaviors during the early breakout of COVID-19: cross-sectional study. J Med Internet Res 2020 Aug 21;22(8):e19995 [FREE Full text] [doi: 10.2196/19995] [Medline: 32716897]
15. Liu N, Kim J, Jung Y, Arisy A, Nicdao MA, Mikaheal M, et al. Remote monitoring systems for chronic patients on home hemodialysis: field test of a copresence-enhanced design. JMIR Hum Factors 2017 Aug 29;4(3):e21 [FREE Full text] [doi: 10.2196/humanfactors.7078] [Medline: 28851680]
16. Zhai Y, Wang Y, Zhang M, Gittell JH, Jiang S, Chen B, et al. From isolation to coordination: how can telemedicine help combat the Covid-19 outbreak? medRxiv. Preprint posted online February 23, 2020. [doi: 10.1101/2020.02.20.20025957]
17. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The role of telehealth in reducing the mental health burden from COVID-19. Telemed J E Health 2020 Apr;26(4):377-379. [doi: 10.1089/tmj.2020.0068] [Medline: 32202977]
18. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
19. Asan O, Cooper Ii F, Nagavally S, Walker RJ, Williams JS, Ozieh MN, et al. Preferences for health information technologies among US adults: analysis of the Health Information National Trends Survey. J Med Internet Res 2018 Oct 18;20(10):e277 [FREE Full text] [doi: 10.2196/jmir.9436] [Medline: 30341048]
20. Jaffe DH, Lee L, Huynh S, Haskell TP. Health inequalities in the use of telehealth in the United States in the lens of COVID-19. Popul Health Manag 2020 Oct;23(5):368-377. [doi: 10.1089/pop.2020.0186] [Medline: 32816644]
21. Li P, Liu X, Mason E, Hu G, Zhou Y, Li W, et al. How telemedicine integrated into China's anti-COVID-19 strategies: case from a National Referral Center. BMJ Health Care Inform 2020 Aug;27(3):e100164. [doi: 10.1136/bmjhci-2020-100164] [Medline: 32830110]
22. Huang X, Wei F, Hu L, Wen L, Chen K. Epidemiology and clinical characteristics of COVID-19. Arch Iran Med 2020 Apr 01;23(4):268-271. [doi: 10.34172/aim.2020.09] [Medline: 32271601]
23. Hu S, Wang W, Wang Y, Litvinova M, Luo K, Ren L, et al. Infectivity, susceptibility, and risk factors associated with SARS-CoV-2 transmission under intensive contact tracing in Hunan, China. medRxiv. Preprint posted online July 24, 2020. [doi: 10.1101/2020.07.23.20160317] [Medline: 32793929]
24. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
25. Nouri S, Khoong EC, Lyles CR, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catalyst 2020 May 04;1(3):1-13 [FREE Full text]
26. Katzow MW, Steinway C, Jan S. Telemedicine and Health Disparities During COVID-19. Pediatrics 2020 Aug;146(2):e20201586. [doi: 10.1542/peds.2020-1586] [Medline: 32747592]

Abbreviations

CDC: Centers for Disease Control and Prevention
GED: General Educational Development
HCP: health care provider

Edited by G Fagherazzi; submitted 23.08.20; peer-reviewed by L Larsen, Z Li; comments to author 26.09.20; revised version received 17.10.20; accepted 01.02.21; published 18.02.21.

Please cite as:
Horrell LN, Hayes S, Herbert LB, MacTurk K, Lawhon L, Valle CG, Bhowmick A
Telemedicine Use and Health-Related Concerns of Patients With Chronic Conditions During COVID-19: Survey of Members of Online Health Communities
J Med Internet Res 2021;23(2):e23795
URL: http://www.jmir.org/2021/2/e23795/
doi:10.2196/23795
PMID:33539307

Original Paper

Impact of the COVID-19 Pandemic on the Psychological Distress of Medical Students in Japan: Cross-sectional Survey Study

Yoshito Nishimura[1], MD, MPH, PhD; Kanako Ochi[1], MD, PhD; Kazuki Tokumasu[1], MD; Mikako Obika[1], MD, PhD; Hideharu Hagiya[1], MD, PhD; Hitomi Kataoka[1], MD, PhD; Fumio Otsuka[1], MD, PhD

Department of General Medicine, Okayama University Hospital, Okayama, Japan

Corresponding Author:
Yoshito Nishimura, MD, MPH, PhD
Department of General Medicine
Okayama University Hospital
2-5-1 Shikata-Cho, Kita-Ku
Okayama
Japan
Phone: 81 86 235 7342
Email: nishimura-yoshito@okayama-u.ac.jp

Abstract

Background: The COVID-19 pandemic has negatively affected medical education. However, little data are available about medical students' distress during the pandemic.

Objective: This study aimed to provide details on how medical students have been affected by the pandemic.

Methods: A cross-sectional study was conducted. A total of 717 medical students participated in the web-based survey. The survey included questions about how the participants' mental status had changed from before to after the Japanese nationwide state of emergency (SOE).

Results: Out of 717 medical students, 473 (66.0%) participated in the study. In total, 29.8% (141/473) of the students reported concerns about the shift toward online education, mostly because they thought online education would be ineffective compared with in-person learning. The participants' subjective mental health status significantly worsened after the SOE was lifted (P<.001). Those who had concerns about a shift toward online education had higher odds of having generalized anxiety and being depressed (odds ratio [OR] 1.97, 95% CI 1.19-3.28) as did those who said they would request food aid (OR 1.99, 95% CI 1.16-3.44) and mental health care resources (OR 3.56, 95% CI 2.07-6.15).

Conclusions: Given our findings, the sudden shift to online education might have overwhelmed medical students. Thus, we recommend that educators inform learners that online learning is not inferior to in-person learning, which could attenuate potential depression and anxiety.

(J Med Internet Res 2021;23(2):e25232) doi:10.2196/25232

KEYWORDS

COVID-19; online education; depression; pandemic; anxiety; medical student

Introduction

The COVID-19 global pandemic has drastically changed our lives, with more than 93 million cases and 2 million deaths reported globally as of January 17, 2021, according to statistics from the World Health Organization [1]. The pandemic led to significant declines in the global economy, and the uncertainty and fears associated with COVID-19 led to increases in mental health disorders. According to a systematic review by Xiong and colleagues, high rates of anxiety, depression, posttraumatic stress disorder, and psychological distress were reported in the general population [2]. Among the people affected, health care workers (HCWs) are considered to be more vulnerable to these issues. A previous cross-sectional study performed in the Asian-Pacific region looked into the psychological impact of COVID-19 on HCWs and noted that the burden of COVID-19 cases might not be related to psychological adversity; however, the increasing number of cases and mortality of COVID-19 continues to threaten the well-being of HCWs [3]. In Japan, more than 300,000 cases have been confirmed as of January 17, 2021, with an explosion in the number of cases in early April, August, and November 2020 [4]. While Japan previously

brought the outbreak under control through active cluster tracing, restriction of mass gatherings, and advocating universal masking and hand hygiene, the country has been experiencing a surge in the number of cases since November 2020, which suggests there were no simple solutions to this issue [5]. After the surge in COVID-19 cases, on April 16, 2020, the Government of Japan declared a state of emergency (SOE) in all 47 prefectures, which lasted until May 25, 2020 [6]. In Okayama, a prefecture in the western part of mainland Japan [7] with approximately 1.9 million people and ranked 20 out of 47 by population among the 47 prefectures, only 2071 confirmed cases had been reported by January 17, 2021. Behind the scenes of successful COVID-19 mitigation, however, those in education-related jobs and medical students struggled with the rapid change in the educational system, as shown in examples from Australia and Spain [8,9].

The SOE was lifted on May 14, 2020, in Okayama based on the low incidence of COVID-19. Okayama University School of Medicine (OUSM), one of the largest national universities in Japan, requested its students to stay home to prevent the spread of COVID-19 within the medical school and the hospital even before the SOE (ie, March 25, 2020). Until the stay-home request was lifted on May 22, 2020, medical students were required to cope with this sudden change in their lifestyle and shifted to online education amid the fear and considerable uncertainty surrounding COVID-19. In the early 2000s, the SARS-CoV outbreak had a devastating impact on academic education, including sudden curriculum changes and rapid integration of information technology [10-12]. Similarly, the current COVID-19 pandemic has provoked significant turmoil in society. In particular, as briefly noted above, mental health problems due to the pandemic have drawn attention worldwide as studies have suggested the need for mental health care interventions during the outbreak [13-18]. Medical students, who essentially need clinical exposure, may have been impacted even further by the pandemic because of the cancellation of clinical rotation. Due to the pandemic that forced educational institutions to eliminate in-person teaching sessions, medical students needed to adapt to new educational environments, such as distance or remote e-learning [19]. While some researchers argued that the COVID-19 pandemic could be an opportunity to catalyze changes in medical education [20], the rapid change in the system and environment might cause significant stress to medical students. Even before the COVID-19 pandemic, the global prevalence of anxiety among medical students was estimated to be 33.8%, and it was the most prevalent in those from the Middle East and Asia [21]. Also, a meta-analysis done in 2016 showed that depression was prevalent in 28.0% of medical students globally. Given the significant psychological distress related to the pandemic [22], the current prevalence of anxiety and depression among medical students might be even higher.

In Japan, medical schools follow an undergraduate medical education system, which is similar to most countries. Students typically enter medical school immediately after high school graduation, often at 18 years old, and they go through 6 years of medical education before graduation. The fundamental philosophy of academic medicine is to provide a quality educational experience. To date, few published studies have investigated the living environment and mental health status of medical students during the COVID-19 pandemic [23-25]. To address this, we conducted a total population survey of OUSM medical students to comprehensively clarify what they require and how they have been mentally affected by COVID-19.

Methods

Study Design, Setting, and Participants

We performed a cross-sectional study that employed an anonymous, self-administered, voluntary web-based survey. Participants included medical students in all years of study at OUSM. We invited all 717 medical students who belonged to the OUSM as of April 1, 2020 (ie, the first day of the academic year in Japan), to participate in the survey. The participants' consent was implied by the completion of the survey.

Our research team developed the survey through consultation with a medical education expert panel at OUSM and through piloting. We conducted a pretest of all the instruments in a sample group of 15 recent medical school graduates, who graduated in April 2020, to confirm their comprehension of the test and to make sure the questions were appropriate to measure psychological distress amid the pandemic. Cronbach α for the instrument developed by the team was .73. Content validity of the survey, as confirmed by the medical education expert panel, showed that the instrument represented the proposed domains (ie, depression and anxiety distress) to be measured in medical students. The survey was administered with Qualtrics (Qualtrics International Inc), a web-based survey platform. We provided survey instructions and instruments in Japanese. We distributed survey links to the students using OUSM official mailing lists. All participants were invited to complete the survey within 1 week (ie, June 8-14, 2020, in Japan Standard Time). No financial incentives were provided for their participation in the survey. The survey included entries on demographics (ie, age, gender, education before entering medical school, employment status on the date of response, changes in employment status due to the COVID-19 pandemic, marital status, living environment, household size, and comorbidities) as well as COVID-19-related items (eg, chance of contracting COVID-19 during the current pandemic), self-learning-associated activities (eg, average amount of time spent self-learning per day), validated depression and anxiety scale instruments, and financial situations. Self-learning is defined as "proactive processes that students use to acquire academic skill" [26]. In this study, the concept of self-learning by participants includes reading medical books, watching webinars, or preparing for shelf exams by themselves. To protect participants' anonymity as much as possible, participants were not prompted to enter their year of study.

Measurements

COVID-19-Related Questions

To evaluate the extent of participants' concerns and preparedness for COVID-19, they were asked the questions listed in Textbox 1.

Textbox 1. Survey questions related to COVID-19.

COVID-19-related questions:

- What do you think the chances are that you will contract COVID-19 during the current pandemic?
- What is your degree of concern about the health status of your family?
- Do you think you have enough information about the symptoms of COVID-19?
- Do you think you have enough information about prevention and treatment of COVID-19?
- Do you feel worried about COVID-19?

To further assess students' concerns, participants were prompted to respond to the following statements:

- I am concerned because my future career formation may be negatively affected due to the COVID-19 pandemic.
- I am concerned because the COVID-19 pandemic may attenuate our relationship to teachers.
- I am concerned because of the disruption to ongoing research or extracurricular activities.
- I am concerned about the shift toward online education.

All questions were evaluated on a 5-point Likert scale, except for the item on the health status of the family, which was evaluated on a 3-point Likert scale. Those who responded with *very concerned* or *concerned* regarding the entry on concern about the shift toward online education were prompted to provide reasons. To describe participants' needs, they were asked to note the types of support they wish to receive from the university if there is a resurgence of COVID-19. These responses were mandatory.

Self-Learning and Related Activities

As noted above, we developed the survey through consultation with a medical education expert panel at our facility. We assumed that several factors might be related to changes in medical students' subjective mental health status. To evaluate subjective mental health status and the average amount of time per day that participants stayed at home, read books, played video games, and learned by themselves, respondents were prompted to answer the following questions, as pertaining to before the SOE order (ie, April 16, 2020) and during the last 2 weeks, with the base date of when participants completed the survey: How many hours a day did you stay at home? How many hours a day did you read books? How many hours a day did you play video games? and How many hours a day did you self-learn?

Depression and Anxiety Disorders

We assessed the presence of depression using the 9-item Patient Health Questionnaire (PHQ-9), a common screening tool for mood disorders. We used the validated Japanese translation of the scale [27]. The total score for the PHQ-9 ranges from 0 to 27, and we defined scores of 10 or more as having *depression*. We screened for anxiety disorders using the Japanese version of the 7-item Generalized Anxiety Disorder scale (GAD-7), which was validated in 2010 [28]. The total score ranges from 0 to 21, and we defined scores of 10 or more as having *anxiety*. We used cut points of 10 or greater in this study because total scores of 10 or more in the PHQ-9 or GAD-7 represent moderate depression or anxiety, respectively. Both instruments ask respondents about their mental health status during the last 2 weeks.

Financial Situation

In Japan, university students typically live on a monthly allowance from parents. According to the latest statistics by the National Federation of University Co-operative Associations, students receive 72,810 Japanese yen (JPY) (approximately US $680) a month on average [29]. Participants were prompted to give their monthly allowance from the following options: *None*, *<30,000 JPY* (US $280), *30,000-49,999 JPY* (US $280-$467), *50,000-69,999 JPY* (US $467-$654), *70,000-99,999 JPY* (US $654-$935), and *≥100,000 JPY* (≥US $935). Respondents were also asked to answer whether they were receiving a scholarship or student loan.

Statistical Analysis

We analyzed the data using JMP statistical software, version 13.1.0 (SAS Institute Inc). We used the Wilcoxon signed-rank test to examine differences in the amount of time participants spent on self-learning-related activities based on nonnormal distribution. For associations between categorical variables containing small sample sizes, we employed the Fisher exact test. To examine the predictive factors of categorical dependent variables, we used univariate logistic regression analyses. The threshold for significance was defined as $P<.05$.

Ethical Approval

This study protocol was approved by the Institutional Review Board of Okayama University Hospital (reference No. 2006-029; approved on June 5, 2020).

Data Availability Statement

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Results

Overview

The response rate to the survey was 66.0%, as 473 out of 717 OUSM students in all 6 years combined completed the survey. Participants' demographic characteristics are summarized in Table 1. Of note, out of 473 respondents, 250 (52.9%) reported that they were engaged in part-time work, while 44 (9.3%)

reported having resigned or lost their jobs due to the COVID-19 pandemic. Out of 473 respondents, 8 (1.7%) and 6 (1.3%) noted that they had a past medical history of anxiety disorders and depression, respectively.

Table 1. Demographic characteristics of the study participants.

Characteristic	Value (N=473)
Age in years (n=471)[a], mean (SD), 95% CI	22.0 (3.3), 21.7-22.3
Gender, n (%)	
Female	161 (34.0)
Male	311 (65.8)
Nonconforming	1 (0.2)
Education before medical school, n (%)	
High school	434 (91.8)
Career college	1 (0.2)
4-year university	31 (6.6)
Master's degree	7 (1.5)
Employment status, n (%)	
Part-time job	250 (52.9)
Schoolwork only: no extra work	221 (46.7)
Self-owned business	2 (0.4)
Resigned or fired due to the pandemic	44 (9.3)
Marital status, n (%)	
Single	467 (98.7)
Married	5 (1.1)
Missing data	1 (0.2)
Living environment, n (%)	
Alone	308 (65.1)
With family	127 (26.8)
With a partner	20 (4.2)
Alone: family or relatives nearby	18 (3.8)
Household size, mean (SD), 95% CI	1.7 (1.3), 1.6-1.8
Comorbidity, n (%)	
None	401 (84.8)
Asthma	37 (7.8)
Anxiety disorder	8 (1.7)
Depression	6 (1.3)
Other	34 (7.2)

[a]Of the 473 respondents, 2 did not specify their age.

COVID-19-Related Survey Items

Table 2 and Figure 1 show respondents' answers to the COVID-19-related survey items. Out of 473 respondents, 81 (17.1%) responded that they were either *likely* or *very likely* to contract COVID-19 during the ongoing pandemic; 275 (58.1%) *agreed* or *strongly agreed* to the prompt "I feel worried about COVID-19." Regarding the breakdown of students' concerns about COVID-19, 182 (38.5%), 121 (25.6%), and 235 (49.7%) out of 473 respondents acknowledged that they were concerned about the negative impacts of COVID-19 on their future career formation, relationship with teachers, and ongoing research or extracurricular activities, respectively, while 141 (29.8%) also reported concerns about a shift toward online education. The reasons for these concerns included the belief that online education may not be as effective as on-site education (92/141, 65.2%), possible resurgence of the COVID-19 outbreak leading to a sudden change in the curriculum (74/141, 52.5%), and decreased clinical exposure (92/141, 65.2%).

Of the 473 participants, 298 (63.0%) answered that they would request financial aid if a stay-home order recurred due to a COVID-19 resurgence, followed by a request for food aid (100/473, 21.1%), technical support for online education (100/473, 21.1%), and mental health care resources, including counseling by therapists or psychologists (85/473, 18.0%).

Table 2. Results of the COVID-19-related survey items.

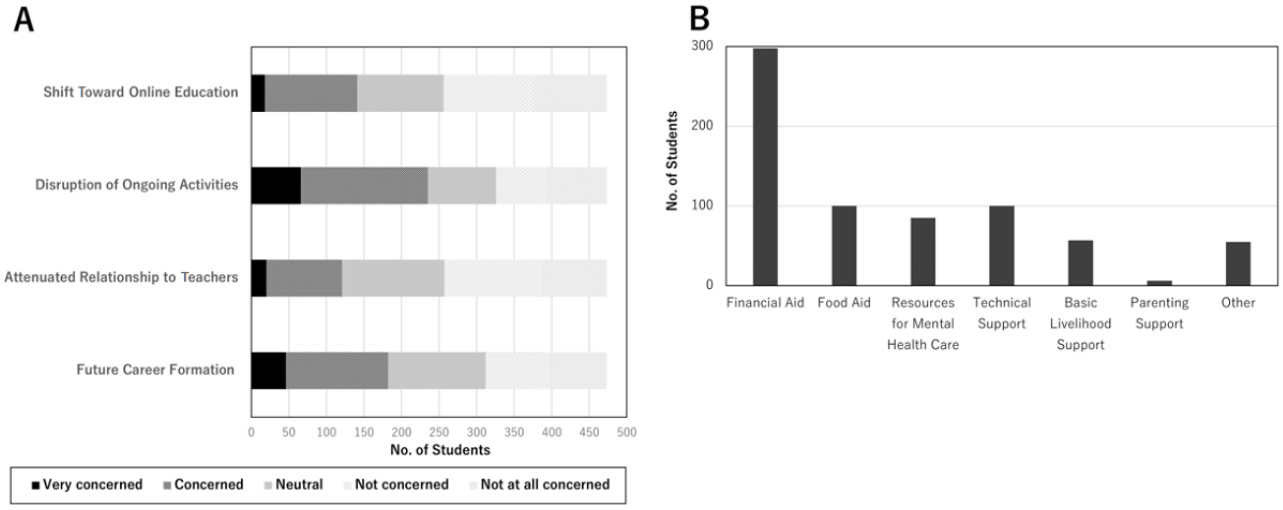
Survey entry and responses	Value (N=473), n (%)
"Chance of contracting COVID-19 during the current pandemic"	
Very likely	15 (3.2)
Likely	66 (14.0)
Neutral	238 (50.3)
Unlikely	121 (25.6)
Very unlikely	33 (6.9)
"Degree of concern about the health status of your family"	
Very concerned	198 (41.9)
Somewhat concerned	237 (50.1)
Not concerned	38 (8.0)
"I have enough knowledge about symptoms of COVID-19"	
Strongly agree	8 (1.7)
Agree	158 (33.4)
Neutral	169 (35.7)
Disagree	116 (24.5)
Strongly disagree	22 (4.7)
"I have enough knowledge about the treatment of COVID-19"	
Strongly agree	6 (1.3)
Agree	65 (13.7)
Neutral	138 (29.2)
Disagree	196 (41.4)
Strongly disagree	68 (14.4)
"I feel worried about COVID-19"	
Strongly agree	46 (9.7)
Agree	229 (48.4)
Neutral	106 (22.4)
Disagree	77 (16.3)
Strongly disagree	15 (3.2)
"Concerned about future career formation due to COVID-19"	
Very concerned	46 (9.7)
Concerned	136 (28.8)
Neutral	130 (27.5)
Not concerned	85 (18.0)
Never concerned	76 (16.0)
"Concerned about attenuated relationship to teachers"	
Very concerned	20 (4.2)
Concerned	101 (21.4)
Neutral	136 (28.7)
Not concerned	129 (27.3)
Never concerned	87 (18.4)
"Concerned about disruption of ongoing research or extracurricular activities"	
Very concerned	66 (14.0)

Survey entry and responses	Value (N=473), n (%)
Concerned	169 (35.7)
Neutral	91 (19.2)
Not concerned	67 (14.2)
Never concerned	80 (16.9)
"Concerned about shift toward online education"	
Very concerned	18 (3.8)
Concerned	123 (26.0)
Neutral	115 (24.3)
Not concerned	121 (25.6)
Never concerned	96 (20.3)
"The reasons for concernabout a shift toward online education" (n=141)[a]	
Online education may not be as effective as on-site education	92 (65.2)
Fear of sudden change in the curriculum	74 (52.5)
Less clinical exposure	92 (65.2)
"What supports do you want from the university?"[b]	
Financial aid (eg, tuition waiver and no-interest student loan)	298 (63.0)
Food aid	100 (21.1)
Resources for mental health care	85 (18.0)
Technical support (eg, free rental of personal computer and pocket Wi-Fi)	100 (21.1)
Basic livelihood support (eg, resources for health maintenance or preventing domestic violence)	57 (12.1)

[a]This survey entry was explicitly answered by participants who answered *very concerned* or *concerned* in response to the item "Concerned about shift toward online education." Multiple answers were permitted.

[b]Multiple answers were permitted.

Figure 1. (A) The breakdown of medical students' educational concerns related to COVID-19. (B) What medical students want from the university upon the second wave the COVID-19 pandemic.



Change in Self-Learning-Related Amount of Time Before and After the SOE

We compared the change in the average amount of time per day that respondents spent at home, reading books, playing video games, and self-learning before the SOE and within 2 weeks prior to the survey completion, after the SOE. As shown in Table 3 and Figure S1 in Multimedia Appendix 1, the participants spent significantly longer on all the activities mentioned above after the SOE compared to before the SOE (P<.001). There were also significant differences in their subjective mental health status before and after the SOE, based on the Wilcoxon signed-rank test (P<.001).

Table 3. Comparison of age-weighted self-learning and related activity times of medical students before and after the national state of emergency (SOE) in Japan.

Measure	Before SOE (N=473)	Last 2 weeks[a] (N=473)	*P* value[b]
Activity (hours/day), mean (95% CI)			
Stay at home	16.1 (15.7-16.6)	18.8 (18.5-19.2)	<.001
Reading books	0.70 (0.58-0.82)	1.1 (0.96-1.2)	<.001
Playing video games	1.5 (1.3-1.7)	2.1 (1.9-2.3)	<.001
Self-learning	2.6 (2.4-2.8)	4.0 (3.7-4.2)	<.001
Subjective mental health status, n (%)			
Very good	85 (18.0)	39 (8.2)	<.001
Good	171 (36.1)	58 (12.2)	<.001
Fair	193 (40.8)	286 (60.5)	<.001
Bad	18 (3.8)	76 (16.1)	<.001
Very bad	6 (1.3)	14 (3.0)	<.001

[a]The base date of the *last 2 weeks* is the date when the participants answered the survey. The survey period was from June 8 to 14, 2020.

[b]*P* values were calculated with the Wilcoxon signed-rank test.

Regression Analyses of Factors Associated With Depression and Anxiety

Of the 473 participants, 75 (15.9%) had PHQ-9 scores of 10 or more and 34 (7.2%) had GAD-7 scores of 10 or more. Table 4 presents all the results from the univariate regression analyses. The odds of being depressed were significantly higher in those who had concerns about a shift toward online education (odds ratio [OR] 1.97, 95% CI 1.19-3.28) and in those who said they would request food aid (OR 1.99, 95% CI 1.16-3.44) and mental health care resources (OR 3.56, 95% CI 2.07-6.15) from the university in the event of the resurgence of COVID-19. Regarding generalized anxiety, the odds were higher in respondents who said they would request food aid (OR 2.50, 95% CI 1.21-5.20) and mental health care resources (OR 3.16, 95% CI 1.51-6.59).

Table 4. Results of univariate regression analyses of factors associated with depression and generalized anxiety in Japanese medical students.

Dependent variable and independent variables	Odds ratio (95% CI)[a]	*P* value
Depression (n=75), PHQ-9[b] score ≥10		
Female vs male	1.08 (0.58-2.01)	.81
Age	1.06 (0.96-1.16)	.45
Concerned about the following[c]:		
Future career formation	1.40 (0.85-2.30)	.18
Attenuated relationship to teachers	1.46 (0.85-2.50)	.17
Disruption of ongoing research or extracurricular activities	0.81 (0.50-1.33)	.41
Shift toward online education	1.97 (1.19-3.28)	.009
Would request the following:		
Financial aid	0.92 (0.55-1.53)	.74
Food aid	1.99 (1.16-3.44)	.01
Mental health care resources	3.56 (2.07-6.15)	<.001
Anxiety (n=34), GAD-7[d] score ≥10		
Female vs male	0.81 (0.37-1.74)	.58
Age	0.96 (0.85-1.09)	.57
Concerned about the following[c]:		
Future career formation	0.75 (0.36-1.58)	.45
Attenuated relationship to teachers	1.43 (0.68-3.03)	.35
Disruption of ongoing research or extracurricular activities	0.53 (0.26-1.09)	.09
Shift toward online education	1.50 (0.73-3.10)	.27
Would request the following:		
Financial aid	1.25 (0.59-2.62)	.56
Food aid	2.50 (1.21-5.20)	.01
Mental health care resources	3.16 (1.51-6.59)	.002

[a]Univariate regression analysis was performed for each dependent variable.

[b]PHQ-9: 9-item Patient Health Questionnaire.

[c]Participants who answered *very concerned* or *concerned* in the respective survey entries listed in Table 2 were considered the *positive concerns group*. On the other hand, those who answered *neutral*, *not concerned*, or *not at all concerned* were considered the *no concerns group* in this analysis.

[d]GAD-7: 7-item Generalized Anxiety Disorder scale.

Prevalence of Financial Hardship

Table 5 and Figure S2 in Multimedia Appendix 1 show the self-reported monthly allowance from parents and situations regarding financial aid. Out of 473 respondents, 332 (70.2%) had a monthly allowance that was lower than the national average (ie, 72,810 JPY; approximately US $680) [29], while 131 (27.7%) received some sort of financial aid for their living and education expenses.

Table 5. Description of the financial situation of Japanese medical students.

Characteristic	Value (N=473), n (%)
Monthly allowance from parents (JPY[a])	
None	127 (26.8)
<30,000	43 (9.1)
30,000-49,999	81 (17.1)
50,000-69,999	81 (17.1)
70,000-99,999	78 (16.5)
≥100,000	63 (13.3)
Financial aid	
None	342 (72.3)
Student loan only	74 (15.6)
Scholarship only	44 (9.3)
Both student loan and scholarship	13 (2.8)

[a]JPY: Japanese yen; according to the Bank of Japan, the exchange rate of 1 US dollar to JPY was 107.39-107.41 on the morning of June 17, 2020.

Discussion

To the best of our knowledge, this study is the first survey of Japanese medical students regarding their life circumstances and challenges due to COVID-19. As of January 17, 2021, Japan has had 307,696 cumulative COVID-19 cases, the second largest number in the World Health Organization Western Pacific region and the 39th largest worldwide [1]. While Okayama has had comparatively fewer cases than other Japanese prefectures, the study results underscore that the pandemic inflicted profound mental health challenges on medical students. We found that approximately 10% of the students with part-time jobs had lost their jobs due to the pandemic. Furthermore, many medical students had concerns regarding their basic life security, demanding support from the university in the form of financial aid, food aid, technical support, and mental health care resources due to the SOE. As shown in previous studies, there may be an increasing prevalence of food insecurity during the pandemic, which might negatively affect the students' mental well-being [30,31]. Regarding students' subjective psychological distress, those who expressed concerns about the rapid shift toward online education and fear around basic life security were more likely to be depressed and anxious after the SOE.

Concerns around future career disruption, attenuated relationships with medical teachers, and disruption of ongoing extracurricular activities were prevalent among the participants, which underlines the considerable uncertainty amid the COVID-19 pandemic (see Table 2). In particular, 63.0% of respondents reported the need for financial aid in the event of a second wave of the pandemic. These data correspond to the fact that more than 70% of participants received no more than the national average monthly allowance from their parents (see Table 5). Contrary to the general notion in Japan that "medical students are financially well-off," many may have experienced financial hardship. Recently, the theory of willpower has gathered public attention. Willpower is defined as the ability to resist short-term temptations to achieve long-term goals or the capacity to override an unwanted thought or impulse [32,33]. As previously studied, the theory is even applicable to students [34]. Financial stability is considered a prerequisite of willpower, which aids appropriate decision making [32]. Combined with previous research findings showing that financial instability could lead to worse mental health outcomes [35,36], educational institutions are expected to develop a strategy to offer financial support to those in need. Although there are no straightforward solutions to address psychological distress amid the pandemic, we encourage educational institutions to provide emergency grant funding and to lobby the national or local government to offer additional funding given the number of students in critical need.

Regarding lifestyle changes and psychological distress, after the SOE was lifted, the participants experienced significantly worse mental health and spent a significantly longer time at home, reading books, playing video games, and learning by themselves (see Table 3) than before the announcement of the SOE. Our results are consistent with prior longitudinal and cross-sectional studies that showed that the COVID-19 pandemic led to a sedentary lifestyle and psychological distress [13,16,17,37]. While longer self-study times and reading books could be the consequences of staying at home longer and might not have detrimental effects on students' well-being, increasing time spent playing video games might be alarming. A previous study suggested that internet addiction is significantly associated with depression and anxiety [38]. As indicated by a previous study, the increase in gaming could be a coping behavior against psychological stress [39].

As for regression analyses of factors associated with depression and anxiety, in our study population, those who said they would request food aid and mental health care resources from the university upon the future resurgence of the COVID-19 outbreak had significantly higher odds of depression and anxiety. Furthermore, concerns around the shift toward online education were identified as factors associated with depression (see Table 4). Surprisingly, 65.2% of those concerned about the shift

toward online education thought online education was less effective than in-person education. While previous studies have reported the utility and noninferiority of online learning compared to offline in-person learning [40,41], this study's results revealed a potential gap in perception regarding the effectiveness of online education between medical students and educators. Educators should not assume that students know the potential benefits of online learning, and it is essential to inform learners that online learning is not inferior to in-person learning, which could attenuate potential depression and anxiety. While in-person communication has become difficult nowadays due to the fear of COVID-19, medical schools may need to reach out to students to find out who among them are suffering from underlying life insecurity and provide multilateral support, including early mental health care interventions. In particular, cognitive behavioral therapy (CBT) has drawn attention as an effective solution to help mitigate psychological distress. Ho et al discussed mental health strategies to support HCWs amid the COVID-19 pandemic, and they noted that CBT could mitigate maladaptive coping behaviors, including self-blame, by enhancing stress-managing skills [42]. In particular, internet CBT (I-CBT) using a learning management platform such as Moodle has been noted as a cost-effective and valid solution to address psychiatric symptoms [43,44]. Further studies are warranted to determine whether I-CBT may also be an effective solution for medical students' psychological distress in Japan.

Although this study's findings might be discouraging, medical students potentially have a role during the COVID-19 pandemic. In several countries including the United States and Vietnam, the mobilization of medical students toward a frontline COVID-19 response has been reported as a considerable help in assisting HCWs while increasing the clinical experience of the students [45,46]. While mobilization of medical students to help in the COVID-19 response needs to be voluntary, this is undoubtedly something that medical students can contribute to the community, health care system, and society. Knowing about good examples from other countries amid the pandemic might help medical students raise their hopes and overcome psychological distress. Educators are encouraged to provide supports to meet individual student needs, and the approach should not be a one-size-fits-all type. A longitudinal nationwide study to follow up with the mental health of medical students is warranted.

Several limitations of this study should be noted. First, due to the single-center, cross-sectional survey design, we may not conclude causal relationships. Second, cross-sectional research contains a limitation in terms of addressing changes over time. While we illustrated the changes in the average amount of time that students spent on self-learning-related activities before and after the SOE in Table 3, a longitudinal study design would be more appropriate to examine the differences in the study cohort. Third, we asked participants to provide their mental health status and amount of time spent on self-learning-related activities before the SOE, approximately 6 weeks before the survey implementation, both of which are subject to recall bias. Fourth, we did not check the convergent and structural validity of the instrument developed by our team. Thus, it should be noted that these scores might not represent hypothesized domains (ie, depression and anxiety). Next, the PHQ-9 and the GAD-7 are self-reported questionnaires that measure psychiatric symptoms, and no clinical diagnoses of depression or anxiety were made. It should be noted that the gold standard for establishing psychiatric diagnosis involves structured clinical interviews and functional neuroimaging [47,48]. Also, PHQ-9 and GAD-7 scores were obtained only after the SOE. Thus, they might not have changed during the period from before the SOE to after. Lastly, due to the nature of the survey topic, those interested in the COVID-19 public health emergency or in mental health may have been more likely to respond, leading to self-selection bias. Despite these limitations, a total population sampling strategy coupled with higher-than-usual response rates [49] contributed to high internal validity.

In conclusion, through this study, we have provided graphical data and evidence regarding the impacts of the COVID-19 pandemic on medical students. In circumstances of considerable uncertainty, both educators and medical students need to be flexible, patient, and resilient. The uncertainty and drastic change triggered substantial psychological distress in students, which was greater than we had assumed. Although Okayama prefecture survived the pandemic with fewer confirmed COVID-19 cases than other prefectures, medical students experienced significant adverse impacts due to the public health emergency. Educational institutions should recognize the prevalence of basic needs insecurity, such as financial difficulties and a shortage of staples, including food. We encourage educational institutions to offer emergency grant funding and lobby the government to provide additional funding at a policy level, given the number of students in critical need. As medical educators, we need to be accountable for the advantages of online education in the field of medicine to alleviate students' psychological distress, in addition to providing multilateral support to those in need, including early mental health care interventions such as I-CBT. While we targeted medical students in a single Japanese national university, the survey results warrant further research and analysis to determine whether the observed distress was amplified by existing anxiety, depression, burnout, etc, or whether it was a wholly COVID-19-related phenomenon. We call for increased research in populations with more COVID-19 cases than in Japan to figure out the challenges that medical students from different cultures and backgrounds have been facing amid the pandemic.

Acknowledgments

We would like to thank the Academic Affairs Division of OUSM for its cooperation in conducting the survey.

Authors' Contributions

YN wrote the manuscript, designed the study, and analyzed the data. KO, KT, MO, HH, and HK analyzed the data and revised the manuscript. FO designed and supervised the research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Amount of time respondents spent on self-learning and related activities and their self-reported financial situation.
[PNG File , 110 KB - jmir_v23i2e25232_app1.png]

References

1. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. 2021. URL: https://covid19.who.int/table [accessed 2021-02-11]
2. Xiong J, Lipsitz O, Nasri F, Lui LMW, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
3. Chew NWS, Ngiam JN, Tan BY, Tham S, Tan CY, Jing M, et al. Asian-Pacific perspective on the psychological well-being of healthcare workers during the evolution of the COVID-19 pandemic. BJPsych Open 2020 Oct 08;6(6):e116 [FREE Full text] [doi: 10.1192/bjo.2020.98] [Medline: 33028449]
4. Novel coronavirus (COVID-19). Ministry of Health, Labour and Welfare Japan. 2021. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000164708_00079.html [accessed 2021-02-11]
5. Looi M. COVID-19: Japan ends state of emergency but warns of "new normal". BMJ 2020 May 26;369:m2100. [doi: 10.1136/bmj.m2100] [Medline: 32457055]
6. Basic Policies for Novel Coronavirus Disease Control by the Government of Japan (Summary). Tokyo, Japan: Government of Japan; 2020 Mar 28. URL: https://www.mhlw.go.jp/content/10900000/000634753.pdf [accessed 2021-02-11]
7. Current population estimates as of October 1, 2019. Statistics Bureau of Japan. 2019. URL: https://www.stat.go.jp/english/data/jinsui/2019np/index.html [accessed 2021-02-11]
8. Azorín C. Beyond COVID-19 supernova. Is another education coming? J Prof Cap Community 2020 Jun 26;5(3/4):381-390. [doi: 10.1108/jpcc-05-2020-0019]
9. Drane CF, Vernon L, O'Shea S. Vulnerable learners in the age of COVID-19: A scoping review. Aust Educ Res 2020 Nov 27:1-20 [FREE Full text] [doi: 10.1007/s13384-020-00409-5] [Medline: 33262554]
10. Clark J. Fear of SARS thwarts medical education in Toronto. BMJ 2003 Apr 12;326(7393):784 [FREE Full text] [doi: 10.1136/bmj.326.7393.784/c] [Medline: 12689971]
11. Patil NG, Chan Y, Yan H. SARS and its effect on medical education in Hong Kong. Med Educ 2003 Dec;37(12):1127-1128 [FREE Full text] [doi: 10.1046/j.1365-2923.2003.01723.x] [Medline: 14984121]
12. Sherbino J, Atzema C. "SARS-Ed": Severe acute respiratory syndrome and the impact on medical education. Ann Emerg Med 2004 Sep;44(3):229-231 [FREE Full text] [doi: 10.1016/j.annemergmed.2004.05.021] [Medline: 15332063]
13. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: Longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
14. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e17-e18 [FREE Full text] [doi: 10.1016/S2215-0366(20)30077-8] [Medline: 32085841]
15. Shigemura J, Kurosawa M. Mental health impact of the COVID-19 pandemic in Japan. Psychol Trauma 2020 Jul;12(5):478-479. [doi: 10.1037/tra0000803] [Medline: 32525392]
16. Talevi D, Socci V, Carai M, Carnaghi G, Faleri S, Trebbi E, et al. Mental health outcomes of the COVID-19 pandemic. Riv Psichiatr 2020;55(3):137-144. [doi: 10.1708/3382.33569] [Medline: 32489190]
17. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
18. Zhang Y, Zhang H, Ma X, Di Q. Mental health problems during the COVID-19 pandemic and the mitigation effects of exercise: A longitudinal study of college students in China. Int J Environ Res Public Health 2020 May 25;17(10):3722 [FREE Full text] [doi: 10.3390/ijerph17103722] [Medline: 32466163]
19. Al-Balas M, Al-Balas HI, Jaber HM, Obeidat K, Al-Balas H, Aborajooh EA, et al. Distance learning in clinical medical education amid COVID-19 pandemic in Jordan: Current situation, challenges, and perspectives. BMC Med Educ 2020 Oct 02;20(1):341 [FREE Full text] [doi: 10.1186/s12909-020-02257-4] [Medline: 33008392]
20. Lucey CR, Johnston SC. The transformational effects of COVID-19 on medical education. JAMA 2020 Sep 15;324(11):1033-1034. [doi: 10.1001/jama.2020.14136] [Medline: 32857137]

21. Quek TTC, Tam WWS, Tran BX, Zhang M, Zhang Z, Ho CSH, et al. The global prevalence of anxiety among medical students: A meta-analysis. Int J Environ Res Public Health 2019 Jul 31;16(15):2735 [FREE Full text] [doi: 10.3390/ijerph16152735] [Medline: 31370266]
22. Puthran R, Zhang MWB, Tam WW, Ho RC. Prevalence of depression amongst medical students: A meta-analysis. Med Educ 2016 Apr;50(4):456-468. [doi: 10.1111/medu.12962] [Medline: 26995484]
23. Guadix SW, Winston GM, Chae JK, Haghdel A, Chen J, Younus I, et al. Medical student concerns relating to neurosurgery education during COVID-19. World Neurosurg 2020 Jul;139:e836-e847 [FREE Full text] [doi: 10.1016/j.wneu.2020.05.090] [Medline: 32426066]
24. Loh TY, Hsiao JL, Shi VY. COVID-19 and its effect on medical student education in dermatology. J Am Acad Dermatol 2020 Aug;83(2):e163-e164 [FREE Full text] [doi: 10.1016/j.jaad.2020.05.026] [Medline: 32413448]
25. Theoret C, Ming X. Our education, our concerns: The impact on medical student education of COVID-19. Med Educ 2020 Jul;54(7):591-592 [FREE Full text] [doi: 10.1111/medu.14181] [Medline: 32310318]
26. Zheng B, Ward A, Stanulis R. Self-regulated learning in a competency-based and flipped learning environment: Learning strategies across achievement levels and years. Med Educ Online 2020 Dec;25(1):1686949 [FREE Full text] [doi: 10.1080/10872981.2019.1686949] [Medline: 31672119]
27. Muramatsu K, Miyaoka H, Kamijima K, Muramatsu Y, Tanaka Y, Hosaka M, et al. Performance of the Japanese version of the Patient Health Questionnaire-9 (J-PHQ-9) for depression in primary care. Gen Hosp Psychiatry 2018;52:64-69. [doi: 10.1016/j.genhosppsych.2018.03.007] [Medline: 29698880]
28. Muramatsu K, Miyaoka H, Ueshima K, Muramatsu Y, Fuse K, Yoshimine H. Validation and utility of a Japanese version of the GAD-7. Jpn J Psychosom Med 2010;50(6):592. [doi: 10.15064/jjpm.50.6_592_2]
29. Summary of campus life data 2019 [Article in Japanese]. National Federation of University Co-operative Associations. 2019. URL: https://www.univcoop.or.jp/press/life/report.html [accessed 2021-02-11]
30. Owens MR, Brito-Silva F, Kirkland T, Moore CE, Davis KE, Patterson MA, et al. Prevalence and social determinants of food insecurity among college students during the COVID-19 pandemic. Nutrients 2020 Aug 20;12(9):2515 [FREE Full text] [doi: 10.3390/nu12092515] [Medline: 32825251]
31. Wolfson JA, Leung CW. Food insecurity and COVID-19: Disparities in early effects for US adults. Nutrients 2020 Jun 02;12(6):1648 [FREE Full text] [doi: 10.3390/nu12061648] [Medline: 32498323]
32. Baumeister RF, Tice DM, Vohs KD. The strength model of self-regulation: Conclusions from the second decade of willpower research. Perspect Psychol Sci 2018 Mar;13(2):141-145. [doi: 10.1177/1745691617716946] [Medline: 29592652]
33. Bechara A. Decision making, impulse control and loss of willpower to resist drugs: A neurocognitive perspective. Nat Neurosci 2005 Nov;8(11):1458-1463. [doi: 10.1038/nn1584] [Medline: 16251988]
34. Duckworth AL, Seligman MEP. Self-discipline outdoes IQ in predicting academic performance of adolescents. Psychol Sci 2005 Dec;16(12):939-944. [doi: 10.1111/j.1467-9280.2005.01641.x] [Medline: 16313657]
35. Richardson T, Elliott P, Roberts R, Jansen M. A longitudinal study of financial difficulties and mental health in a national sample of British undergraduate students. Community Ment Health J 2017 Apr;53(3):344-352 [FREE Full text] [doi: 10.1007/s10597-016-0052-0] [Medline: 27473685]
36. McCloud T, Bann D. Financial stress and mental health among higher education students in the UK up to 2018: Rapid review of evidence. J Epidemiol Community Health 2019 Oct;73(10):977-984 [FREE Full text] [doi: 10.1136/jech-2019-212154] [Medline: 31406015]
37. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
38. Ho RC, Zhang MWB, Tsang TY, Toh AH, Pan F, Lu Y, et al. The association between internet addiction and psychiatric co-morbidity: A meta-analysis. BMC Psychiatry 2014 Jun 20;14:183 [FREE Full text] [doi: 10.1186/1471-244X-14-183] [Medline: 24947851]
39. Balhara YPS, Kattula D, Singh S, Chukkali S, Bhargava R. Impact of lockdown following COVID-19 on the gaming behavior of college students. Indian J Public Health 2020 Jun;64(Supplement):S172-S176 [FREE Full text] [doi: 10.4103/ijph.IJPH_465_20] [Medline: 32496250]
40. Pei L, Wu H. Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis. Med Educ Online 2019 Dec;24(1):1666538 [FREE Full text] [doi: 10.1080/10872981.2019.1666538] [Medline: 31526248]
41. Seymour-Walsh AE, Weber A, Bell A. Practical approaches to pedagogically rich online tutorials in health professions education. Rural Remote Health 2020 May;20(2):6045 [FREE Full text] [doi: 10.22605/RRH6045] [Medline: 32471311]
42. Ho CS, Chee CY, Ho RC. Mental health strategies to combat the psychological impact of COVID-19 beyond paranoia and panic. Ann Acad Med Singap 2020 Mar 16;49(3):155-160 [FREE Full text] [Medline: 32200399]
43. Zhang MWB, Ho RCM. Moodle: The cost effective solution for internet cognitive behavioral therapy (I-CBT) interventions. Technol Health Care 2017;25(1):163-165. [doi: 10.3233/THC-161261] [Medline: 27689560]
44. Soh HL, Ho RC, Ho CS, Tam WW. Efficacy of digital cognitive behavioural therapy for insomnia: A meta-analysis of randomised controlled trials. Sleep Med 2020 Nov;75:315-325. [doi: 10.1016/j.sleep.2020.08.020] [Medline: 32950013]

45. Soled D, Goel S, Barry D, Erfani P, Joseph N, Kochis M, et al. Medical student mobilization during a crisis: Lessons from a COVID-19 medical student response team. Acad Med 2020 Sep;95(9):1384-1387 [FREE Full text] [doi: 10.1097/ACM.0000000000003401] [Medline: 32282373]
46. Tran BX, Vo LH, Phan HT, Pham HQ, Vu GT, Le HT, et al. Mobilizing medical students for COVID-19 responses: Experience of Vietnam. J Glob Health 2020 Dec;10(2):020319 [FREE Full text] [doi: 10.7189/jogh.10.020319] [Medline: 33110521]
47. Ho CSH, Lim LJH, Lim AQ, Chan NHC, Tan RS, Lee SH, et al. Diagnostic and predictive applications of functional near-infrared spectroscopy for major depressive disorder: A systematic review. Front Psychiatry 2020;11:378 [FREE Full text] [doi: 10.3389/fpsyt.2020.00378] [Medline: 32477179]
48. Husain SF, Yu R, Tang T, Tam WW, Tran B, Quek TT, et al. Validating a functional near-infrared spectroscopy diagnostic paradigm for major depressive disorder. Sci Rep 2020 Jun 16;10(1):9740 [FREE Full text] [doi: 10.1038/s41598-020-66784-2] [Medline: 32546704]
49. Cunningham CT, Quan H, Hemmelgarn B, Noseworthy T, Beck CA, Dixon E, et al. Exploring physician specialist response rates to web-based surveys. BMC Med Res Methodol 2015 Apr 09;15:32 [FREE Full text] [doi: 10.1186/s12874-015-0016-z] [Medline: 25888346]

Abbreviations

CBT: cognitive behavioral therapy
GAD-7: 7-item Generalized Anxiety Disorder scale
HCW: health care worker
I-CBT: internet cognitive behavioral therapy
JPY: Japanese yen
OR: odds ratio
OUSM: Okayama University School of Medicine
PHQ-9: 9-item Patient Health Questionnaire
SOE: state of emergency

Edited by G Eysenbach; submitted 23.12.20; peer-reviewed by R Ho, MA Hasdianda, MA Bahrami, H Pratomo; comments to author 15.01.21; revised version received 18.01.21; accepted 31.01.21; published 18.02.21.

Please cite as:
Nishimura Y, Ochi K, Tokumasu K, Obika M, Hagiya H, Kataoka H, Otsuka F
Impact of the COVID-19 Pandemic on the Psychological Distress of Medical Students in Japan: Cross-sectional Survey Study
J Med Internet Res 2021;23(2):e25232
URL: http://www.jmir.org/2021/2/e25232/
doi:10.2196/25232
PMID:33556033

Viewpoint

Precision Assessment of COVID-19 Phenotypes Using Large-Scale Clinic Visit Audio Recordings: Harnessing the Power of Patient Voice

Paul J Barr[1,2], PhD, MScPH; James Ryan[3], DO; Nicholas C Jacobson[2,4,5], PhD

[1]The Dartmouth Institute for Health Policy & Clinical Practice, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States

[2]The Center for Technology and Behavioral Health, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States

[3]Ryan Family Practice, Ludington, MI, United States

[4]Biomedical Data Science, Geisel School of Medicine at Dartmouth College, Lebanon, NH, United States

[5]Department of Psychiatry, Geisel School of Medicine at Dartmouth and Dartmouth Hitchcock Health, Lebanon, NH, United States

Corresponding Author:
Paul J Barr, PhD, MScPH
The Dartmouth Institute for Health Policy & Clinical Practice
Geisel School of Medicine at Dartmouth College
46 Centerra Parkway
Lebanon, NH, 03766
United States
Phone: 1 603 653 0863
Email: paul.j.barr@dartmouth.edu

Abstract

COVID-19 cases are exponentially increasing worldwide; however, its clinical phenotype remains unclear. Natural language processing (NLP) and machine learning approaches may yield key methods to rapidly identify individuals at a high risk of COVID-19 and to understand key symptoms upon clinical manifestation and presentation. Data on such symptoms may not be accurately synthesized into patient records owing to the pressing need to treat patients in overburdened health care settings. In this scenario, clinicians may focus on documenting widely reported symptoms that indicate a confirmed diagnosis of COVID-19, albeit at the expense of infrequently reported symptoms. While NLP solutions can play a key role in generating clinical phenotypes of COVID-19, they are limited by the resulting limitations in data from electronic health records (EHRs). A comprehensive record of clinic visits is required—audio recordings may be the answer. A recording of clinic visits represents a more comprehensive record of patient-reported symptoms. If done at scale, a combination of data from the EHR and recordings of clinic visits can be used to power NLP and machine learning models, thus rapidly generating a clinical phenotype of COVID-19. We propose the generation of a pipeline extending from audio or video recordings of clinic visits to establish a model that factors in clinical symptoms and predict COVID-19 incidence. With vast amounts of available data, we believe that a prediction model can be rapidly developed to promote the accurate screening of individuals at a high risk of COVID-19 and to identify patient characteristics that predict a greater risk of a more severe infection. If clinical encounters are recorded and our NLP model is adequately refined, benchtop virologic findings would be better informed. While clinic visit recordings are not the panacea for this pandemic, they are a low-cost option with many potential benefits, which have recently begun to be explored.

(J Med Internet Res 2021;23(2):e20545) doi:10.2196/20545

KEYWORDS

communication; coronavirus; COVID-19; Machine Learning; natural language processing; patient-physician communication; patient records; recording

Challenges in Identifying COVID-19 Clinical Phenotypes

COVID-19 cases are exponentially increasing worldwide; however, clinical COVID-19 phenotypes remain unclear. A clinical phenotype is an observable characteristic (ie, symptom) of a disease in a particular individual. A meta-analysis of COVID-19–related symptom presentations reported that the most frequent clinical symptoms are fever, cough, fatigue, and dyspnea [1]. However, the meta-analysis reported considerable

heterogeneity (I^2=84.9%-96.4%) among studies, potentially suggestive of the extreme heterogeneity among these symptoms at the individual patient level. Other less frequent COVID-19 symptoms include anosmia, dysgeusia, headache, sore throat, rhinorrhea, diarrhea, nausea, and myalgias [2]. However, their clinical implication, prevalence, and importance remain unclear.

A traditional reductionist approach to identifying COVID-19 treatments is not as simple as extrapolating the current knowledge toward our limited SARS-CoV-2 model. Clinical treatments are often based on a set of established biochemical markers, and reports of less frequent symptoms of a disease may reveal a biochemical pathway that can be subjected to pharmacotherapeutic intervention with previously unreported agents. Only laboratory tests can confirm a diagnosis of COVID-19, but such tests are in short supply. This presents an unprecedented need to develop better assessment methods to identify and generate heterogeneity in the clinical profile of COVID-19 and other viral diseases across the entire health care system. The urgency of this need cannot be understated, as it holds a key to understand how to identify and treat COVID-19 more accurately.

Using "Big Data" to Understand the Clinical Manifestations of COVID-19

Natural language processing (NLP) and machine learning may yield a method to rapidly identify individuals at a high risk for COVID-19 and to understand key symptoms upon clinical manifestation and presentation [3]. The existing applications of NLP and machine learning in medical diagnostics are based on a combination of structured (eg, symptom codes, medications, laboratory findings, etc) and unstructured (eg, visit notes, radiology reports, etc) data recorded by clinicians in patients' electronic health records (EHRs). Using NLP and machine learning approaches, data on documented signs and symptoms in the EHR are already being used to identify clinical conditions (computational phenotyping) [4]. Such NLP-based efforts are currently being applied to unstructured text data captured in the EHR from telehealth consultations to develop better screening tools for COVID-19 [5]. Ancillary data can improve the accuracy of computational phenotyping, such as information from disease registries. However, the performance of any model is determined by the quality of data used to generate it, and concerns exist about the fullness of data captured in the EHR.

Limitations of EHR Data

This considerable degree of symptom heterogeneity reported among patients with COVID-19 can deter the accurate documentation of less frequently reported symptoms in the EHR. Documentation inaccuracies in electronic medical records are not a new phenomenon; an analysis of data from 105 clinics indicated that 90% of clinician notes had at least one error, including 636 documentation errors that accounted for 181 charted findings that did not take place and 455 findings that were not charted [6]. Data on such symptoms may not be accurately synthesized into patient records owing to the pressing need to treat patients in overburdened health care settings. In this scenario, clinicians may focus on documenting widely reported symptoms that suggest a diagnosis of COVID-19 albeit at the expense of infrequently reported symptoms because overburdened clinicians are more likely to be affected by cognitive biases such as anchoring and confirmation biases [7]. Additionally, codes of the International Classification of Diseases (10th revision), the mainstay of documentation in electronic medical records, do not adequately capture COVID-19–related symptoms [8]. While NLP solutions can play a key role in generating clinical phenotypes of COVID-19, they are limited by the resulting limitations in EHR data. A comprehensive record of the clinic visits is required—an audio recording may be the solution [9].

Clinical Phenotypes Based on Audio Recordings of Clinic Visits

A small but growing number of health systems routinely obtain audio recordings, and, in some cases, video recordings of clinic visits [9,10]. For example, human scribes are commonly employed to review recordings of clinic visits and make detailed notes, thus reducing the documentation burden on clinicians and improving the accuracy of data entered in the EHR. A recording of the clinic visit represents a more comprehensive and accurate record of patient-reported symptoms. If performed at scale, a combination of data from the EHR and recordings of clinic visits can be used to power NLP and machine learning models, thus rapidly generating a clinical phenotype of COVID-19 and infections with subsequent SARS-CoV-2 strains. In addition to a more comprehensive record of symptoms discussed, recordings also asynchronously collect additional ancillary information such as the type and frequency of cough, which can help improve the precision of phenotyping.

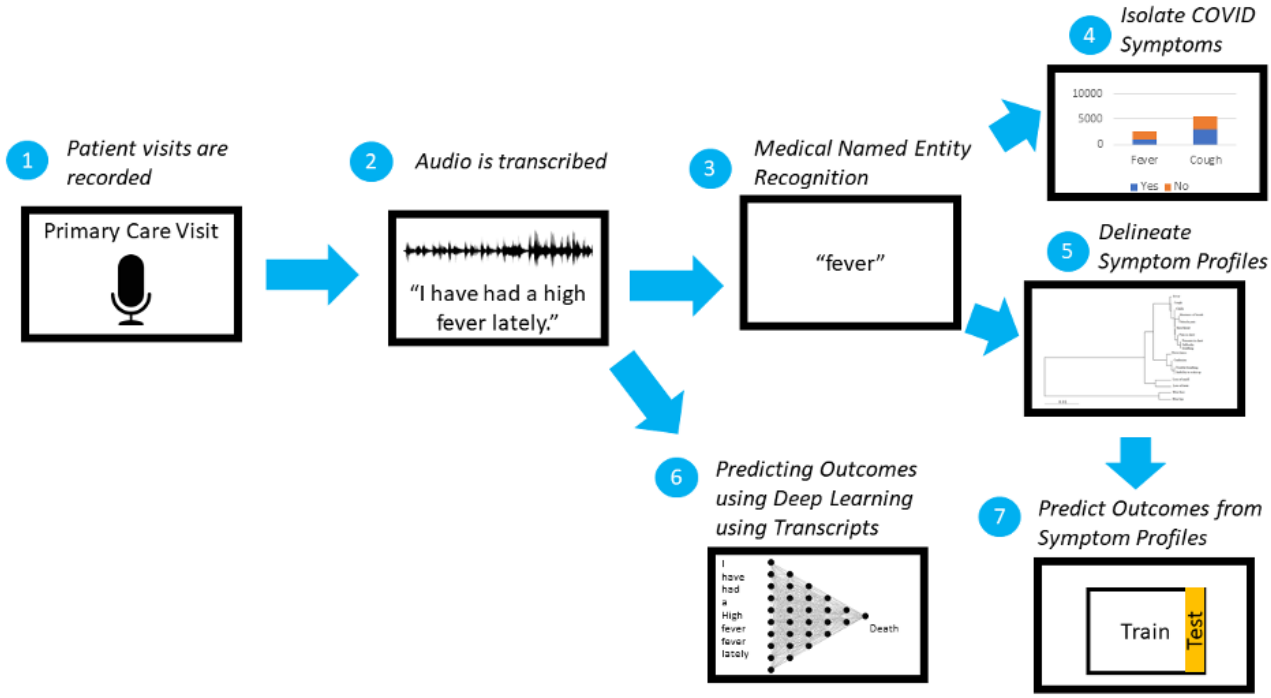
The generation of NLP and machine learning models requires the transcription of vast quantities of conversations of patients being investigated for COVID-19 upon clinic visits (with subsequent confirmatory laboratory tests for the disease) and the annotation of these transcripts by annotators trained to identify symptom mentions. The performance of automated speech recognition algorithms has significantly improved [10], allowing for the real-time use of audio data rather than transcripts of audio data, which are more time-consuming to obtain. Real-time risk assessment is critical when responding to an infectious disease such as COVID-19, since it allows for individuals to identify their risk level and more rapidly self-isolate, thus reducing the risk of disease transmission. Data annotation to generate models that can accurately identify symptoms is not without its challenges, many of which have been summarized by Quiroz et al [11]. It can be difficult for annotators to identify vaguely indicated symptoms from the unstructured natural language used in clinic visit conversations, with a negative impact on model performance. Rigorous training of annotators can help mitigate this challenge; however, such training and annotation is time-consuming and would require a large team of annotators to rapidly meet the immediate need for such an analysis. In addition, model training requires human input and time. Furthermore, the generation of optimal data would require continuous data refinement, wherein records of

suspected cases are replaced by the findings of confirmatory tests so as not to correspond to clinician views or biases.

Implications of the Adoption of Clinic Visit Recordings in Managing COVID-19

We propose the generation of a pipeline from the audio recordings of clinic visits to models based on clinical symptoms and the prediction of COVID-19 incidence (Figure 1). With vast amounts of available data, we believe a prediction model can be rapidly developed to promote accurate screening of individuals at risk of COVID-19. Beyond the challenge of generating a clinical phenotype, an unfiltered account of a patient's clinical experience of the disease allows us to answer other pressing questions, such as those related to understanding the constellation of patient characteristics that may predict a greater risk of a more severe infection. If clinical consultations are recorded and our NLP model is adequately refined, benchtop virologic findings are better informed. Recordings of clinic visits also provide a historic reference, such that we may be better prepared for subsequent pandemics.

Figure 1. Natural language processing pipeline from audio recordings to the establishment of a clinical phenotype of COVID-19.



With the mass transition to telehealth consultations and the availability of guidance for conducting remote assessments of COVID-19 via telehealth at primary care centers [12], an opportunity to capture audio recordings of consultations at scale is now available. An accurate model predicting a higher risk of COVID-19 could be applied to telehealth consultations with the added benefit of reducing the exposure risk among clinicians, patients, and the general public. The use of NLP for remote COVID-19 screening is already emerging; for example, audio recordings of cough sounds are being used to identify individuals with COVID-19 [13,14].

Data From Beyond the Clinic

While recordings of clinic visits are not the panacea for this pandemic, they are a low-cost alternative with many potential benefits that have recently begun to be explored. Beyond audio recordings, video recordings of telehealth consultations can provide additional diagnostic information such as skin appearance [12]. At-home voice-based technologies such as Amazon Alexa, Apple's Siri, and Google Home can also be used, allowing further information from outside of clinic visits to supplement predictive models [15]. For example, the Mayo Clinic has recently added a skill to Amazon Alexa called "Answers on COVID-19," which provides resources on COVID-19 and a virtual questionnaire to determine a person's symptoms and whether the person should get tested for COVID-19 [16].

Considering current accelerated efforts to manage COVID-19, care must be taken to rigorously protect sensitive data, with existing challenges in accessing the corpus of patient recordings needed to generate these models [11]. A data collection method should only be used entirely with an opt-in voluntary framework to preserve privacy and confidentiality; however, this method can help obtain data on COVID-19 symptom exacerbation at a scale unattainable with all traditional methods. This, as is often the case, points toward an evolving learning health system capable of managing computable knowledge.

Acknowledgments

We would like to acknowledge Prof Lisa Marsch for reviewing our manuscript and Dr Jesse Schoonmaker for editing it.

Conflicts of Interest

None declared.

References

1. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. Int J Infect Dis 2020 May;94:91-95 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.017] [Medline: 32173574]
2. McIntosh K, Hirsch M, Bloom A. Coronavirus disease 2019 (COVID-19). UpToDate. URL: https://www.uptodate.com/contents/coronavirus-disease-2019-covid-19-clinical-features [accessed 2021-02-12]
3. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. BMJ 2015 Apr 24;350:h1885 [FREE Full text] [doi: 10.1136/bmj.h1885] [Medline: 25911572]
4. Richesson RL, Hammond WE, Nahm M, Wixted D, Simon GE, Robinson JG, et al. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. J Am Med Inform Assoc 2013 Dec;20(e2):e226-e231 [FREE Full text] [doi: 10.1136/amiajnl-2013-001926] [Medline: 23956018]
5. Obeid J, Davis M, Turner M, Meystre S, Heider P, O'Bryan EC, et al. An artificial intelligence approach to COVID-19 infection risk assessment in virtual visits: A case report. J Am Med Inform Assoc 2020 Aug 01;27(8):1321-1325 [FREE Full text] [doi: 10.1093/jamia/ocaa105] [Medline: 32449766]
6. Weiner S, Wang S, Kelly B, Sharma G, Schwartz A. How accurate is the medical record? A comparison of the physician's note with a concealed audio recording in unannounced standardized patient encounters. J Am Med Inform Assoc 2020 May 01;27(5):770-775. [doi: 10.1093/jamia/ocaa027] [Medline: 32330258]
7. Yuen T, Derenge D, Kalman N. Cognitive bias: Its influence on clinical diagnosis. J Fam Pract 2018 Jun;67(6):366-372. [Medline: 29879236]
8. Crabb B, Lyons A, Bale M, Martin V, Berger B, Mann S, et al. Comparison of International Classification of Diseases and Related Health Problems, Tenth Revision Codes With Electronic Medical Records Among Patients With Symptoms of Coronavirus Disease 2019. JAMA Netw Open 2020 Aug 03;3(8):e2017703 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.17703] [Medline: 32797176]
9. Elwyn G, Barr PJ, Piper S. Digital clinical encounters. BMJ 2018 May 14;361:k2061. [doi: 10.1136/bmj.k2061] [Medline: 29760149]
10. Barr PJ, Bonasia K, Verma K, Dannenberg MD, Yi C, Andrews E, et al. Audio-/Videorecording Clinic Visits for Patient's Personal Use in the United States: Cross-Sectional Survey. J Med Internet Res 2018 Sep 12;20(9):e11308 [FREE Full text] [doi: 10.2196/11308] [Medline: 30209029]
11. Quiroz JC, Laranjo L, Kocaballi AB, Berkovsky S, Rezazadegan D, Coiera E. Challenges of developing a digital scribe to reduce clinical documentation burden. NPJ Digit Med 2019;2:114 [FREE Full text] [doi: 10.1038/s41746-019-0190-1] [Medline: 31799422]
12. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]
13. Tobias M. AI And Medical Diagnostics: Can A Smartphone App Detect Covid-19 From Speech Or A Cough? Forbes. URL: https://www.forbes.com/sites/marcwebertobias/2020/05/05/ai-and-medical-diagnostics-can-a-smartphone-app-detect-covid-19-from-speech-or-a-cough/?sh=46cda2815436 [accessed 2021-12-02]
14. Imran A, Posokhova I, Qureshi HN, Masood U, Riaz MS, Ali K, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Inform Med Unlocked 2020;20:100378 [FREE Full text] [doi: 10.1016/j.imu.2020.100378] [Medline: 32839734]
15. Dojchinovski D, Ilievski A, Gusev M. Interactive home healthcare system with integrated voice assistant. 2019 Presented at: 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO); 20-24 May 2019; Opatija, Croatia. [doi: 10.23919/MIPRO.2019.8756983]
16. Furst J. Mayo Clinic introduces skill for Amazon's Alexa about COVID-19. Mayo Clinic. URL: https://newsnetwork.mayoclinic.org/discussion/mayo-clinic-introduces-skill-for-amazons-alexa-about-covid-19/ [accessed 2021-02-12]

Abbreviations

EHR: electronic health record
NLP: natural language processing

Edited by G Eysenbach, R Kukafka; submitted 03.06.20; peer-reviewed by M Reblin, Z Yin; comments to author 22.08.20; revised version received 22.09.20; accepted 18.01.21; published 19.02.21.

Please cite as:
Barr PJ, Ryan J, Jacobson NC
Precision Assessment of COVID-19 Phenotypes Using Large-Scale Clinic Visit Audio Recordings: Harnessing the Power of Patient Voice
J Med Internet Res 2021;23(2):e20545
URL: http://www.jmir.org/2021/2/e20545/
doi:10.2196/20545
PMID:33556031

Original Paper

Mobile Remote Monitoring of Intimate Partner Violence Among Pregnant Patients During the COVID-19 Shelter-In-Place Order: Quality Improvement Pilot Study

Tamar Krishnamurti[1], PhD; Alexander L Davis[2], PhD; Beth Quinn[3], RNC; Anabel F Castillo[4], PhD; Kelly L Martin[3], RNC; Hyagriv N Simhan[5], MSc, MD

[1]Department of General Internal Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

[2]Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, United States

[3]UPMC Magee-Womens Hospital, Pittsburgh, PA, United States

[4]Naima Health LLC, Pittsburgh, PA, United States

[5]Department of Obstetrics, Gynecology, and Reproductive Sciences, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

Corresponding Author:
Tamar Krishnamurti, PhD
Department of General Internal Medicine
University of Pittsburgh School of Medicine
200 Meyran Avenue
Parkvale Building Suite 200
Pittsburgh, PA, 15213
United States
Phone: 1 4126924855
Email: tamark@pitt.edu

Abstract

Background: Intimate partner violence (IPV) is one of the leading causes of pregnancy-related death. Prenatal health care providers can offer critical screening and support to pregnant people who experience IPV. During the COVID-19 shelter-in-place order, mobile apps may offer such people the opportunity to continue receiving screening and support services.

Objective: We aimed to examine cases of IPV that were reported on a prenatal care app before and during the implementation of COVID-19 shelter-in-place mandates.

Methods: The number of patients who underwent voluntary IPV screening and the incidence rate of IPV were determined by using a prenatal care app that was disseminated to patients from a single, large health care system. We compared the IPV screening frequencies and IPV incidence rates of patients who started using the app before the COVID-19 shelter-in-place order, to those of patients who started using the app during the shelter-in-place order.

Results: We found 552 patients who started using the app within 60 days prior to the enforcement of the shelter-in-place order, and 407 patients who used the app at the start of shelter-in-place enforcement until the order was lifted. The incidence rates of voluntary IPV screening for new app users during the two time periods were similar (before sheltering in place: 252/552, 46%; during sheltering in place: 163/407, 40%). The overall use of the IPV screening tool increased during the shelter-in-place order. A slight, nonsignificant increase in the incidence of physical, sexual, and psychological violence during the shelter-in-place order was found across all app users (P=.56). Notably, none of the patients who screened positively for IPV had mentions of IPV in their medical charts.

Conclusions: App-based screening for IPV is feasible during times when in-person access to health care providers is limited. Our results suggest that the incidence of IPV slightly increased during the shelter-in-place order. App-based screening may also address the needs of those who are unwilling or unable to share their IPV experiences with their health care provider.

(J Med Internet Res 2021;23(2):e22790) doi:10.2196/22790

KEYWORDS

COVID-19; social isolation, sheltering-in-place; intimate partner violence; domestic violence; pregnancy; telemedicine; telehealth

Introduction

Intimate partner violence (IPV) is one of the leading causes of pregnancy-related death [1-3]. In a study of pregnant women who were enrolled in the Nurse Family Partnerships program, 4.7% of participants reported that they experienced IPV during the first 36 weeks of their pregnancy [4]. The documented prevalence of IPV among pregnant women in the United States ranges from <1% to almost 20%, depending on the type of violence and the source of the estimate [5,6]. IPV prevalence estimates using hospital-based samples tend to be higher than population-based studies. This may be due to the higher number of documentation opportunities or the type of screening. Under shelter-in-place mandates, people who experience violence and abuse face restricted access to protective networks, such as health care professionals. Although increases in the number of IPV-related hotline calls have been reported in the media [7,8] and documented in scientific literature [9,10], there is still limited data on the increased number of IPV cases that have been reported to health care systems during the COVID-19 pandemic. Mobile apps that address IPV have been successfully designed [11]. Therefore, such apps may be effective sources of screening and support during the pandemic [12]. During enforced isolation, social networks are restricted. As such, mobile apps may also support the dissemination of IPV risk information to the public. In this study, we examined IPV cases that were reported through a prenatal care app that was prescribed to patients from a large academic health care system. We specifically examined cases from a single US county that were reported 2 months before and after the enforcement of shelter-in-place mandates in 2020.

Methods

Recruitment

In late September 2019, a large academic health system started prescribing the MyHealthyPregnancy app (Apple v.1.4.7 and Android v.1.8) [13] to pregnant patients during their first prenatal appointment. Prior to the COVID-19 pandemic, this app was poised to launch across the entire health care system as a quality improvement initiative. The MyHealthyPregnancy app uses machine learning algorithms to analyze patient-entered data, model each patient's likelihood of experiencing adverse pregnancy events (eg, hypertension and premature delivery), and assess patients' psychosocial risks. The app also offers relevant resources (eg, local women's shelters) to users and notifies their providers about specific risk information in real time (ie, risk information that patients have shared on the app). Furthermore, all app users are notified that their care provider may not see or respond to all risk notifications. The app provides information on immediate actions that users can take to minimize their risks (eg, calling 911, calling their prenatal care provider, or watchfully waiting), based on the seriousness of the identified risks.

All app content was developed in conjunction with a clinical education team that was employed by the health care system. The same team reviewed all app content. The MyHealthyPregnancy app is a product of Naima Health LLC (limited liability company).

The internal protocol for prescribing the MyHealthyPregnancy app involved sending an invitation link to patients' phones and prompting patients to use the app as part of their routine prenatal care. The link allowed users to access a unique code for downloading the app from the Android or Apple app store. By virtue of owning a smartphone, downloading the app, and completing the onboarding process, all users were considered smartphone and internet literate. An individual could not download the app and undergo the onboarding process unless they received a text message–based prescription for the app from their health care provider's office.

The app users included in this study were pregnant residents of Allegheny County, Pennsylvania who were prescribed the MyHealthyPregnancy app during an in-person visit. All app users were offered an opportunity to undergo IPV screening through the app (ie, users received in-app messages that stated "Are you concerned about your safety? Take the pregnancy safety quiz"). Screening was not mandatory, and the screening tool could be accessed from a specific part of the app. We examined data that were collected from January 23 to May 15, 2020, to analyze app-based IPV disclosure before and during the enforcement of local shelter-in-place mandates (ie, approximately 2 months before and after mandate enforcement).

App users consented to the sharing of identifiable data for research purposes and the publication of anonymized aggregate data. Patients did not receive any financial compensation for app use, which was considered a part of routine prenatal care. All analyses were approved by the UPMC health system's quality index review board.

Safety Protocols

App access was protected by a password that was set by the user. The app included a password reset function that required users to input their user ID to receive a personalized text message that provided instructions on how to complete the process. A technical troubleshooting hotline was also available at all times. The IPV screening tool and screening-related notifications used language that was not specific to partner relationships or safety. These notifications were designed to be simple to dismiss, in case there were times when users felt uncomfortable with answering the screening questions. Furthermore, IPV screening data could not be stored or accessed in the app after the screening process was completed, thereby minimizing app users' risk of privacy violations. An icon, which was displayed next to questions about IPV, informed users that positive IPV screening results would be sent to their health care provider. All positive IPV screening results were routinely sent to a UPMC clinical support team.

Measures

The app onboarding process involved answering questions about clinical history (eg, prior preterm births and nulliparity) and demographics. Voluntary, validated screening measures were offered to users after they completed the onboarding process. We used two questions from the Centers for Disease Control and Prevention Behavioral Risk Factor Surveillance System as

measures of physical violence and forced sexual acts. We also used 10 questions from the Women's Experience with Battering scale to quantify psychological abuse [14,15]. Once shelter-in-place mandates were enforced, MyHealthyPregnancy app users were sent a text message that prompted them to use the app to share information about their COVID-19–related protective actions (eg, social distancing) with their care provider. All app users were also offered a COVID-19 triaging tool for assessing their symptoms.

IPV screening is part of the health care system's standard prenatal screening battery. We reviewed the medical charts of patients who screened positively for IPV, to determine whether positive IPV screening tests were documented and to identify all in-person interactions that took place within the system (ie, emergency room, routine prenatal care, specialist care, and physical therapy appointment interactions) during the course of patients' pregnancies.

Statistical Analysis

We compared the IPV screening frequencies and IPV incidence rates of patients who started using the app before the shelter-in-place order, to those of patients who started using the app during the shelter-in-place order, by conducting a two-sample Chi-square test of proportions. IPV incidence rates were then queried against patients' medical charts, to determine whether IPV was documented by a health care provider. Analyses were conducted with R version 3.6.0 (R Foundation for Statistical Computing).

Results

In total, 552 users completed the MyHealthyPregnancy app onboarding process within 60 days prior to the enforcement of shelter-in-place orders (ie, January 23 to March 22, 2020), whereas 407 users completed the onboarding process between the start and end of shelter-in-place enforcement (ie, March 23 to May 15, 2020). Of the 284 respondents who answered all the questions about COVID-19 protective actions, approximately 281 (99%) reported that they adhered to shelter-in-place measures. Frequently performed protective actions included washing hands with soap and water multiple times per day; using hand sanitizer multiple times per day; wearing a face mask; and avoiding public spaces, gatherings, or crowds (eg, not socializing with people who had a high risk of SARS-CoV-2 infection or not frequenting restaurants). Table 1 shows the demographics of patients who completed the onboarding process before and during the shelter-in-place order.

Table 1. MyHealthyPregnancy app user demographics. The demographic distribution of the sample reflects the demographic distribution of the health care system's pregnant population.

Variables	Completed onboarding before the shelter-in-place order (N=552), n (%)	Completed onboarding during the shelter-in-place order (N=407), n (%)
Race/ethnicity		
Black/African American	62 (11)	40 (10)
White	430 (78)	330 (81)
Hispanic/Latinx	13 (2)	3 (1)
Other	45 (8)	33 (8)
Missing data	2 (<1)	1 (<1)
Education status		
No high school or General Education Development diploma	24 (4)	13 (3)
High school or General Education Development diploma	162 (29)	125 (31)
Associate degree	58 (11)	56 (14)
Bachelor's degree	155 (28)	93 (23)
Postgraduate	150 (27)	117 (29)
Missing data	3 (1)	3 (1)
Relationship Status		
Single	20 (4)	21 (5)
Ongoing relationship	159 (29)	115 (28)
Married	369 (67)	269 (66)
Divorced/separated	4 (1)	1 (<1)
Missing data	0 (0)	1 (<1)

The number of patients who used the in-app IPV risk assessment tool did not differ significantly between patients who completed the onboarding process before the shelter-in-place order (252/552, 46%), and those who completed the onboarding

process during the shelter-in-place order (163/407, 40%; two-sample Chi-square test of proportions: 95% CI −12% to −0.01%; *P*=.10). However, the app use rate of patients who completed the onboarding process before the shelter-in-place order was slightly lower than that of patients who completed the onboarding process during the shelter-in-place order. Moreover, the use of the in-app IPV risk assessment across all app users increased from 67% (368/552) to 85% (347/407) during the shelter-in-place order (95% CI 17%-28%; *P*<.001).

Patients who completed the onboarding process during the shelter-in-place order reported that they experienced similar levels of physical violence, sexual violence, and psychological abuse before and during the shelter-in-place order (Table 2). However, after considering all patients who had access to the app during the two time periods (Table 2), we observed a slight, but nonsignificant increase in the incidence of all forms of violence (*P*=.56). Notably, none of the physically at-risk patients (ie, those identified by the app) had any mentions of IPV in their medical chart. However, 24% (4/17) of physically at-risk patients received emergency room care, and 29% (5/17) underwent physical therapy or a prenatal consultation for nonspecific pain or injury.

Table 2. Reports of intimate partner violence and the results of intimate partner violence screening.

Type of intimate partner violence	Before shelter-in-place order			During shelter-in-place order		
	New unique screens vs new unique patients, n/N (%)	New unique positive reports vs new unique patients, n/N (%)	All positive reports vs all reports, n/N (%)	New unique screens vs new unique patients, n/N (%)	New unique positive reports vs new unique patients, n/N (%)	All positive reports vs all reports, n/N (%)
Physical violence[a]	252/552 (46)	2/552 (0.4)	4/461 (0.87)	163/407 (40)	2/407 (0.5)	6/443 (1.4)
Sexual violence[b]	252/552 (46)	2/552 (0.4)	3/461 (0.65)	163/407 (40)	1/407 (0.2)	4/443 (0.9)
Psychological abuse[c]	252/552 (46)	6/552 (1)	6/461 (1.3)	163/407 (40)	3/407 (0.7)	6/442 (1.4)

[a]A 1-item measure (ie, “Has an intimate partner or ex-partner, hit, slapped, kicked, choked, or otherwise physically hurt you in the past month?”).

[b]A 1-item measure (ie, “Has an intimate partner or ex-partner hit, coerced or forced you into sexual activity against your will in the past month?”).

[c]A 10-item measure based on the Women’s Experience with Battering scale (eg, “I try not to rock the boat because I am afraid of what my partner might do”).

Discussion

Principal Results

In this study, we analyzed a stable, prenatal care app–based IPV self-screening tool that was used during shelter-in-place conditions. During the shelter-in-place order, there was a slight, but nonsignificant increase in the incidence of all forms of IPV (*P*=.56). This finding suggests that during times of social isolation (ie, the COVID-19 pandemic), people continue to use technology to disclose their concerns about an increased risk of IPV to care providers [16,17]. However, care providers should take into account the unique needs of pregnant women, to provide them with opportunities for mitigating pregnancy-related risks. It should be noted that even though we observed a nonsignificant, increasing trend in IPV incidence (*P*=.56), new app users’ engagement with the IPV risk assessment tool was stable during both time periods, while overall use went up. One explanation for this is that during the COVID-19 pandemic, people have limited opportunities to seek help outside of home. Therefore, people may be delaying seeking help until abuse starts to escalate.

Given that the people who screened positively for IPV by using an app-based tool were not documented in routine, in-person screening, we believe that the MyHealthyPregnancy app may address the needs of patients who are unable or prefer not to disclose their IPV experiences directly to their care provider. App-based screening may serve as a complementary form of care for those with limited access to health care providers or other social networks. Such technologies may offer patients an additional method for communicating with care providers during times of limited mobility and care access.

Limitations

IPV is serious, and even though it occurs frequently, IPV is rarely reported, even during routine, in-person screening with trusted health care providers. Although our results suggest that there was an increase in the incidence of IPV during the shelter-in-place order, our sample size was not large enough to detect statistically significant differences. This is most likely due to the low rates of disclosure among pregnant women who experience IPV. Although we hope that app-based screening will provide patients with an additional layer of support (ie, by making screening and appropriate resources available at any time), the decision to undergo screening and seek resources is voluntary, and many people may not feel safe or ready to disclose their IPV-related experiences, especially if they are sheltering in place with an abusive partner.

Another limitation of this study is that we were unable to determine whether app-users who screened positively for IPV actually used the resources that were offered through the app. However, we are currently working on a warm handoff system for providing resources to patients who screen positively for IPV and wish to connect with such resources. Although the app users in our study reflect the general population of patients in our health system, it is possible that people who choose not to use the MyHealthyPregnancy app are more likely to experience IPV. This would result in less effective screening. However, we cannot discern this based on our data. It is also possible that app users who experienced IPV, but chose not to disclose it,

reviewed or accessed the universally available support resources that the MyHealthyPregnancy app offers. We hope that this was the case.

Comparison With Prior Work

Regardless of social distancing mandates, the fact that none of the app-detected IPV cases were documented during emergency room visits or other clinical visits highlights a gap in prenatal care [18] that can be filled by implementing app-based technology methods. After shelter-in-place restrictions are lifted, the transition to telemedicine protocols may result in more instances of remote routine screening. Although our results only suggest that pregnant women are more willing to disclose IPV experiences through an app than they are during an in-person encounter, offering app-based IPV risk screening and appropriate resources is one way to address the needs of pregnant people who experience violence during and beyond times of social isolation. Moreover, app-based screening may offer an additional layer of support to patients who would benefit from universally available resources, but do not receive them, even after they present with physical signs of abuse during in-person care visits. This method of screening also allows care providers to anonymously identify the common signs of abuse and violence that patients may present with when seeking care [19] but are missed during in-person visits.

In this study, the number of patients who screened positively for IPV before and during the shelter-in-place period reflects the lower end of national estimates for IPV incidence rates. Therefore, it is unlikely that we identified all instances of IPV among app users (ie, it is possible is that patients who experience IPV do not frequently use smartphones or apps). As such, it is important to take into account that app-based screening alone does not necessarily identify all instances of IPV, especially if an abusive partner controls smartphone use. Our results also suggest that app-based screening may capture a different set of patients who are at risk of IPV, which means that an app-based approach can be used to complement in-person assessments. An app-based approach also provides patients (ie, those who feel that they may be at risk of IPV) with the freedom to engage with screening tools when they feel safe and comfortable.

Acknowledgments

The time that TK spent on preparing the manuscript was supported, in part, by the National Institutes of Health (grant number: KL2 TR001856). The time that AFC spent on preparing the manuscript was supported, in part, by a grant from the Centers for Disease Control and Prevention (grant number: R44DP006417). The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or Centers for Disease Control and Prevention. The funders had no role in the design and conduct of this study; the collection, management, analysis, and interpretation of the data; the preparation, review, and approval of the manuscript; or the decision to submit the manuscript for publication. We thank Miriam Bernard for assisting with data collection. We thank the University of Pittsburgh Center for Behavioral Health, Media, and Technology for its support.

Conflicts of Interest

Naima Health LLC provided the data for this study. The authors did not receive any financial or material compensation for conducting this study. AFC is a full-time employee of Naima Health LLC. TK, ALD, and HNS are cofounders and equity holders of Naima Health LLC, but they did not receive compensation for conducting this study or disseminating the MyHealthyPregnancy app.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1277 KB - jmir_v23i2e22790_app1.pdf]

References

1. Horon IL, Cheng D. Enhanced surveillance for pregnancy-associated mortality--Maryland, 1993-1998. JAMA 2001 Mar 21;285(11):1455-1459. [doi: 10.1001/jama.285.11.1455] [Medline: 11255421]
2. Wallace ME, Crear-Perry J, Mehta PK, Theall KP. Homicide during pregnancy and the postpartum period in Louisiana, 2016-2017. JAMA Pediatr 2020 Apr 01;174(4):387-388. [doi: 10.1001/jamapediatrics.2019.5853] [Medline: 32011644]
3. Martin SL, Mackie L, Kupper LL, Buescher PA, Moracco KE. Physical abuse of women before, during, and after pregnancy. JAMA 2001 Mar 28;285(12):1581-1584. [doi: 10.1001/jama.285.12.1581] [Medline: 11268265]
4. Scribano PV, Stevens J, Kaizar E, NFP-IPV Research Team. The effects of intimate partner violence before, during, and after pregnancy in nurse visited first time mothers. Matern Child Health J 2013 Feb;17(2):307-318. [doi: 10.1007/s10995-012-0986-y] [Medline: 22426619]
5. Gazmararian JA, Lazorick S, Spitz AM, Ballard TJ, Saltzman LE, Marks JS. Prevalence of violence against pregnant women. JAMA 1996 Jun 26;275(24):1915-1920. [Medline: 8648873]
6. Brownridge DA, Taillieu TL, Tyler KA, Tiwari A, Chan KL, Santos SC. Pregnancy and intimate partner violence: risk factors, severity, and health effects. Violence Against Women 2011 Jul;17(7):858-881. [doi: 10.1177/1077801211412547] [Medline: 21775311]

7. Taub A. A New Covid-19 crisis: Domestic abuse rises worldwide. The New York Times. 2020 Apr 06. URL: https://www.nytimes.com/2020/04/06/world/coronavirus-domestic-violence.html [accessed 2021-02-04]
8. Taub A, Bradley J. As domestic abuse rises, U.K. failings leave victims in peril. The New York Times. 2020 Jul 02. URL: https://www.nytimes.com/interactive/2020/07/02/world/europe/uk-coronavirus-domestic-abuse.html [accessed 2021-02-04]
9. Boserup B, McKenney M, Elkbuli A. Alarming trends in US domestic violence during the COVID-19 pandemic. Am J Emerg Med 2020 Dec;38(12):2753-2755 [FREE Full text] [doi: 10.1016/j.ajem.2020.04.077] [Medline: 32402499]
10. Bradbury-Jones C, Isham L. The pandemic paradox: The consequences of COVID-19 on domestic violence. J Clin Nurs 2020 Jul;29(13-14):2047-2049 [FREE Full text] [doi: 10.1111/jocn.15296] [Medline: 32281158]
11. Glass NE, Perrin NA, Hanson GC, Bloom TL, Messing JT, Clough AS, et al. The longitudinal impact of an internet safety decision aid for abused women. Am J Prev Med 2017 May;52(5):606-615. [doi: 10.1016/j.amepre.2016.12.014] [Medline: 28108189]
12. Emezue C. Digital or digitally delivered responses to domestic and intimate partner violence during COVID-19. JMIR Public Health Surveill 2020 Jul 30;6(3):e19831 [FREE Full text] [doi: 10.2196/19831] [Medline: 32678797]
13. Krishnamurti T, Davis AL, Wong-Parodi G, Fischhoff B, Sadovsky Y, Simhan HN. Development and testing of the MyHealthyPregnancy app: A behavioral decision research-based tool for assessing and communicating pregnancy risk. JMIR Mhealth Uhealth 2017 Apr 10;5(4):e42 [FREE Full text] [doi: 10.2196/mhealth.7036] [Medline: 28396302]
14. Chang JC, Cluss PA, Burke JG, Hawker L, Dado D, Goldstrohm S, et al. Partner violence screening in mental health. Gen Hosp Psychiatry 2011;33(1):58-65. [doi: 10.1016/j.genhosppsych.2010.11.009] [Medline: 21353129]
15. Smith PH, Smith JB, Earp JAL. Beyond the measurement trap: A reconstructed conceptualization and measurement of woman battering. Psychol Women Q 1999 Mar 01;23(1):177-193. [doi: 10.1111/j.1471-6402.1999.tb00350.x]
16. Evans ML, Lindauer M, Farrell ME. A pandemic within a pandemic - Intimate partner violence during Covid-19. N Engl J Med 2020 Dec 10;383(24):2302-2304. [doi: 10.1056/NEJMp2024046] [Medline: 32937063]
17. Matoori S, Khurana B, Balcom MC, Koh DM, Froehlich JM, Janssen S, et al. Intimate partner violence crisis in the COVID-19 pandemic: how can radiologists make a difference? Eur Radiol 2020 Dec;30(12):6933-6936 [FREE Full text] [doi: 10.1007/s00330-020-07043-w] [Medline: 32607631]
18. O'Doherty L, Hegarty K, Ramsay J, Davidson LL, Feder G, Taft A. Screening women for intimate partner violence in healthcare settings. Cochrane Database Syst Rev 2015 Jul 22;2015(7):CD007007 [FREE Full text] [doi: 10.1002/14651858.CD007007.pub3] [Medline: 26200817]
19. Miller E, McCaw B, Humphreys BL, Mitchell C. Integrating intimate partner violence assessment and intervention into healthcare in the United States: a systems approach. J Womens Health (Larchmt) 2015 Jan;24(1):92-99 [FREE Full text] [doi: 10.1089/jwh.2014.4870] [Medline: 25606823]

Abbreviations

IPV: intimate partner violence
LLC: limited liability company

Edited by C Basch; submitted 24.07.20; peer-reviewed by C Emezue, E van der Velde; comments to author 21.09.20; revised version received 01.10.20; accepted 31.01.21; published 19.02.21.

Please cite as:
Krishnamurti T, Davis AL, Quinn B, Castillo AF, Martin KL, Simhan HN
Mobile Remote Monitoring of Intimate Partner Violence Among Pregnant Patients During the COVID-19 Shelter-In-Place Order: Quality Improvement Pilot Study
J Med Internet Res 2021;23(2):e22790
URL: http://www.jmir.org/2021/2/e22790/
doi:10.2196/22790
PMID:33605898

Original Paper

Prognosis Score System to Predict Survival for COVID-19 Cases: a Korean Nationwide Cohort Study

Sung-Yeon Cho[1,2*], MD, PhD; Sung-Soo Park[1,3*], MD, PhD; Min-Kyu Song[4,5], MD; Young Yi Bae[1], RN; Dong-Gun Lee[1,2*], MD, PhD; Dong-Wook Kim[1,3*], MD, PhD

[1]Catholic Hematology Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[2]Division of Infectious Diseases, Department of Internal Medicine, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[3]Division of Hematology, Department of Internal Medicine, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

[4]Data Research Institute, YMDtech Inc, Seoul, Republic of Korea

[5]St. Mary's Gong-Gam Mental Health Clinic, Siheung-si, Gyeonggi-do, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Dong-Gun Lee, MD, PhD
Catholic Hematology Hospital, College of Medicine
The Catholic University of Korea
Seoul St. Mary's Hospital
222 Banpo-daero, Seocho-Gu
Seoul, 06591
Republic of Korea
Phone: 82 222586003
Email: symonlee@catholic.ac.kr

Abstract

Background: As the COVID-19 pandemic continues, an initial risk-adapted allocation is crucial for managing medical resources and providing intensive care.

Objective: In this study, we aimed to identify factors that predict the overall survival rate for COVID-19 cases and develop a COVID-19 prognosis score (COPS) system based on these factors. In addition, disease severity and the length of hospital stay for patients with COVID-19 were analyzed.

Methods: We retrospectively analyzed a nationwide cohort of laboratory-confirmed COVID-19 cases between January and April 2020 in Korea. The cohort was split randomly into a development cohort and a validation cohort with a 2:1 ratio. In the development cohort (n=3729), we tried to identify factors associated with overall survival and develop a scoring system to predict the overall survival rate by using parameters identified by the Cox proportional hazard regression model with bootstrapping methods. In the validation cohort (n=1865), we evaluated the prediction accuracy using the area under the receiver operating characteristic curve. The score of each variable in the COPS system was rounded off following the log-scaled conversion of the adjusted hazard ratio.

Results: Among the 5594 patients included in this analysis, 234 (4.2%) died after receiving a COVID-19 diagnosis. In the development cohort, six parameters were significantly related to poor overall survival: older age, dementia, chronic renal failure, dyspnea, mental disturbance, and absolute lymphocyte count <1000/mm^3. The following risk groups were formed: low-risk (score 0-2), intermediate-risk (score 3), high-risk (score 4), and very high-risk (score 5-7) groups. The COPS system yielded an area under the curve value of 0.918 for predicting the 14-day survival rate and 0.896 for predicting the 28-day survival rate in the validation cohort. Using the COPS system, 28-day survival rates were discriminatively estimated at 99.8%, 95.4%, 82.3%, and 55.1% in the low-risk, intermediate-risk, high-risk, and very high-risk groups, respectively, of the total cohort (P<.001). The length of hospital stay and disease severity were directly associated with overall survival (P<.001), and the hospital stay duration was significantly longer among survivors (mean 26.1, SD 10.7 days) than among nonsurvivors (mean 15.6, SD 13.3 days).

Conclusions: The newly developed predictive COPS system may assist in making risk-adapted decisions for the allocation of medical resources, including intensive care, during the COVID-19 pandemic.

(J Med Internet Res 2021;23(2):e26257) doi:10.2196/26257

KEYWORDS

COVID-19; length of stay; mortality; prognosis; triage; digital health; prediction; cohort; risk; allocation; disease management; intensive care; decision making

Introduction

Since the outbreak of unexplained pneumonia in Wuhan, China, in December 2019, which was subsequently identified as COVID-19 caused by the newly discovered pathogen SARS-CoV-2, the COVID-19 pandemic remains active in over 180 countries [1,2]. Globally, as of November 6, 2020, 1,231,017 of a total of 48,534,508 patients with COVID-19 have died, representing an overall infection fatality rate of 2.54% [3,4]. The clinical spectrum of COVID-19 includes asymptomatic or presymptomatic, upper and lower respiratory tract infections and acute respiratory distress syndrome [5-8]. Although the majority of viral infections are self-limiting, COVID-19 cases that are severe (ie, dyspnea, hypoxia, or >50% of lung involvement observed on imaging within 24–48 h) or critical (ie, respiratory failure, shock, or multiple organ failure) are of global concern and require medical resources for intensive care. The proportion of severe or critical COVID-19 cases and the corresponding case fatality rates vary by region and country, ranging from 10%-30% [8-10] to 2%-10%, respectively [3,4].

Most severe or critical cases occur in older patients or those with underlying comorbidities such as cancer, chronic obstructive pulmonary diseases, heart failure, or diabetes [11-13]. The clinical course of patients with COVID-19 also depends on multiple factors, including the immune status of the host, viral load of SARS-CoV-2, genetic diversity of the virus, and underlying diseases. However, details of viral factors and the host immune status (eg, cytokines released) are difficult to analyze in a real-world setting. Therefore, prognosis prediction systems should comprise basic factors such as initial symptoms at diagnosis, vital signs, hemogram parameters, and major underlying comorbidities.

As the global pandemic continues, the ability to detect, in a timely manner, patients with COVID-19 who are at a high risk of death and provide them with intensive care is important. Accordingly, the assessment of disease severity or mortality probability can be used to establish a sustainable strategy. Therefore, in this study, we aimed to develop an easy and simple scoring system that can predict COVID-19 mortality according to the patient's initial presentation and several major underlying comorbidities. Additionally, we investigated the association between the length of hospital stay and disease severity and survival status of patients with COVID-19.

Methods

Study Design and Data Source

This was a nationwide, retrospective cohort study on COVID-19 cases in Korea. For the purpose of this study, COVID-19 cases were defined based on laboratory confirmation of infection and positive results of SARS-CoV-2 reverse transcription polymerase chain reaction assays performed using the testing kits approved by the Korea Ministry of Food and Drug Safety, irrespective of the patient's clinical signs and symptoms [14,15]. Clinical data of patients in this cohort were managed by the Korea Disease Control and Prevention Agency (KDCA) and disclosed to the researchers after application and consent for research purposes in July 2020. The clinical and epidemiological information thus obtained included data of 5628 COVID-19 cases collected between January and April 2020. KDCA is an organization that aims to protect the Korean population from diseases, including emerging infectious diseases such as COVID-19, through national surveillance, health care research, and promotion of policies regarding disease prevention management. Patient data collected for this study included demographic and epidemiological characteristics, hemogram parameters at admission, maximal severity, and clinical outcome obtained from designated hospitals. Patients in the final cohort were randomly allocated to two subcohorts by using a random number generator: two-thirds into the "development cohort" and the remaining one-third into the "validation cohort." The predictive score was developed based on the development cohort, whereas the power of prediction was explored in the validation cohort.

This study was approved by the institutional review board of Seoul St. Mary's Hospital, Seoul, Korea (KC20ZADI0654). Individual patient consent was waived because the data retrieved were anonymous and publicly available.

COVID-19 Management Setting

In Korea, all suspected or confirmed cases of COVID-19 must be reported to the KDCA, as COVID-19 is regarded as a notifiable infectious disease. As a result, all patients with laboratory-confirmed COVID-19 were admitted to designated hospitals or residential treatment centers for isolation, monitoring of symptoms, and treatment. Clinical severity was classified into the following 8 levels according to patient performance, oxygen requirement, and organ failure [16]: (1) no limit of activity, (2) limited activity without oxygen supplementation, (3) requirement of oxygen supply with nasal cannula, (4) requirement of oxygen supply with facial mask, (5) requirement of noninvasive ventilation, (6) requirement of invasive ventilation, (7) requirement of extracorporeal membrane oxygenation (ECMO) for multiple organ failure, or (8) death. Death outcomes were evaluated regardless of maximum severity. Patients that required oxygen supply with invasive ventilation or ECMO were considered as invasive intensive care cases.

Statistical Analysis

Normally distributed numerical variables are presented as mean and SD values. Categorical variables are shown as absolute numbers and their proportions (n, %). The hospital stay durations between two independent groups were compared using Student *t* test. Overall survival was defined as the time from COVID-19 diagnosis to death due to any cause or up to the date of the last follow-up. Death events were censored at the time of hospital discharge for patients who were discharged. Overall survival

rates at 14 days and 28 days were calculated using the Kaplan-Meier method and compared using the Log-rank test.

Within the development cohort, all risk factors with a *P* value <.05 in the univariable analysis were entered into the multivariable model to identify factors associated with overall survival. Multivariable analysis was performed using the Cox proportional hazard regression model. We identified potential variables for the final prediction model based on 2000 bootstrap sampled datasets. When a parameter occurred in 60% or more of the bootstrap models, it was evaluated in the final multiple logistics regression model. We then computed the hazard ratios, 95% CIs, and *P* values for all metrics of the bootstrapped datasets in the final regression model. The final parameters used in the scoring system were defined by a *P* value <.05 in the final regression model. To confirm the risk score for each significant parameter, we adjusted the hazard ratio values to a $\log_e$ scale, followed by the conversion of the respective $\log_e$ scale to a rounded integer point. In the validation cohort, the area under the receiver operating characteristic (AUROC) curve was measured to evaluate the prediction accuracy of the survival rates after 14 and 28 days. An AUROC value above 0.8 was considered reliable. Among the developed risk groups, we compared the length of hospital stay using one-way analysis of variance. For all statistical analyses, we used R statistical software (ver. 3.6.1, R Foundation for Statistical Computing). Statistical significance was set at *P*<.05.

Results

Patient Characteristics

In total, 5628 confirmed COVID-19 cases were reported between January and April 2020. Cases with a postmortem diagnosis (n=7) or lack of clinical course data after diagnosis (n=27) were excluded from the analysis. As shown in Figure 1, a total of 5594 patients with COVID-19 were included in this cohort. Overall, 41.2% (2307/5594) of the patients were male, and 52.2% (2919/5594) were aged 50 years or older. Baseline demographics are summarized in Table 1. In the total cohort, the most frequent age distribution was 50-59 years (1140/5594, 20.4%), followed by 20-29 years (1110/5594, 19.8%) and 60-69 years (905/5594, 16.2%). The most frequently reported symptoms included sputum (1610/5594, 28.8%), fever (1300/5594, 23.2%), and dyspnea (662/5594, 11.8%). Common underlying comorbidities reported were hypertension (1196/5594, 21.4%) and diabetes (686/5594, 12.3%). Moreover, 4% (224/5594) of the patients had dementia and 3.2% (179/5594) of them had cardiac diseases. Distribution of these variables between the development (n=3729) and validation (n=1865) cohorts is presented in Table 1.

Figure 1. Flow diagram of the nationwide cohort of patients with COVID-19.

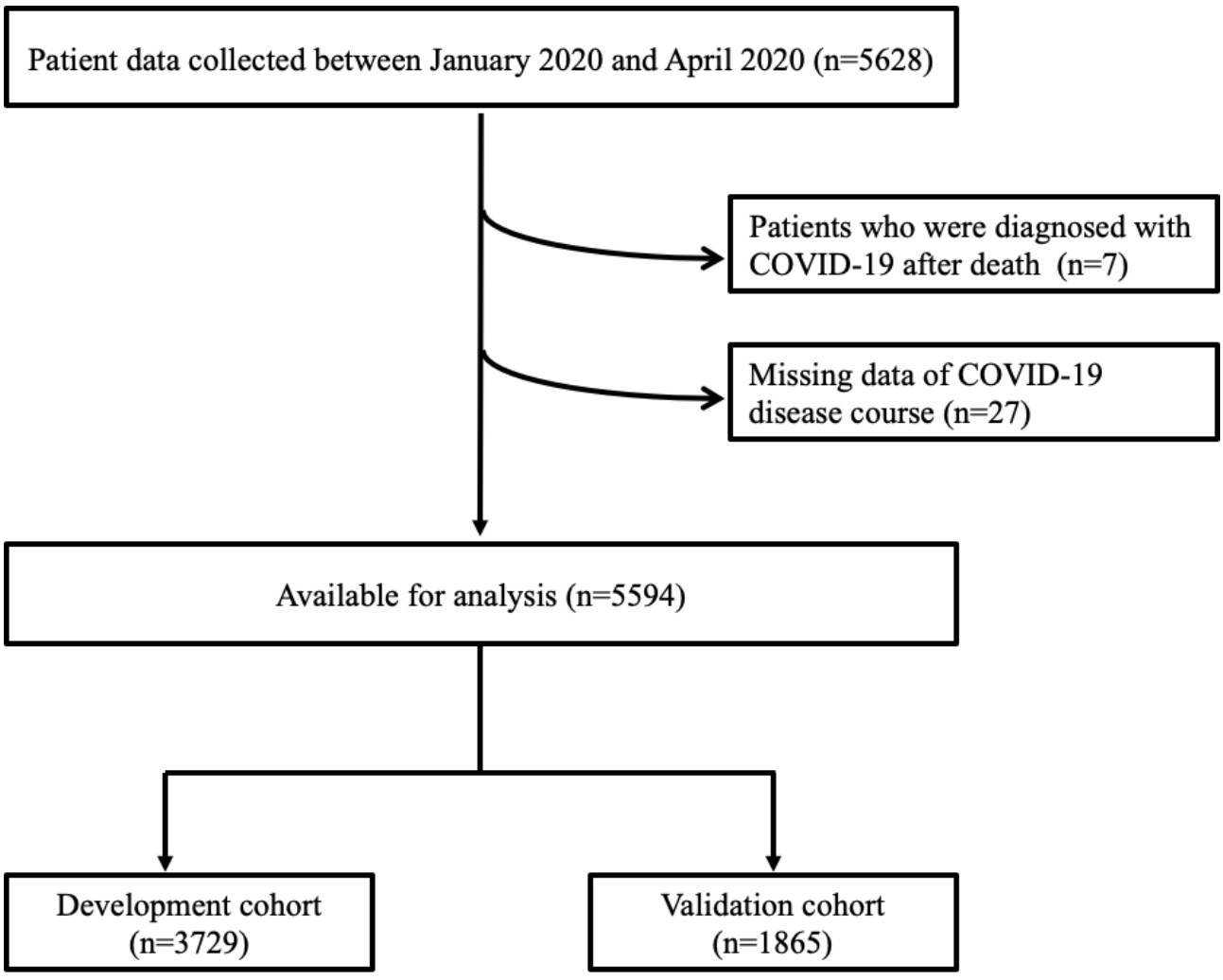


Table 1. Demographics of the study cohorts. All values presented in the table represent data collected at the time of COVID-19 diagnosis.

Variable	Total cohort (N=5594), n (%)	Subcohorts	
		Development cohort (n=3729), n (%)	Validation cohort (n=1865), n (%)
Demographics, n (%)			
Age (years)			
0-9	66 (1.2)	46 (1.2)	20 (1.1)
10-19	205 (3.7)	146 (3.9)	59 (3.2)
20-29	1,110 (19.8)	725 (19.4)	385 (20.6)
30-39	564 (10.1)	378 (10.1)	186 (10)
40-49	739 (13.2)	487 (13.1)	252 (13.5)
50-59	1,140 (20.4)	768 (20.6)	372 (19.9)
60-69	905 (16.2)	605 (16.2)	300 (16.1)
70-79	542 (9.7)	364 (9.8)	178 (9.5)
≥ 80	323 (5.8)	210 (5.6)	113 (6.1)
Gender (male)	2,307 (41.2)	1536 (41.2)	1094 (58.7)
Comorbidity, n (%)			
Hypertension (missing n=3)	1196 (21.4)	795 (21.3)	401 (21.5)
Diabetes (missing n=3)	686 (12.3)	452 (12.1)	234 (12.5)
Dementia (missing n=329)	224 (4)	150 (4)	74 (4)
Cardiac disease (missing n=19)	179 (3.2)	122 (3.3)	57 (3.1)
Cancer in active treatment[a] (missing n=4)	143 (2.6)	90 (2.4)	53 (2.8)
Asthma (missing n=3)	128 (2.3)	82 (2.2)	46 (2.5)
Chronic hepatic disease[b] (missing n=326)	82 (1.5)	47 (1.3)	35 (1.9)
Heart failure (missing n=3)	58 (1)	36 (1)	22 (1.2)
Chronic renal failure (missing n=3)	55 (1)	36 (1)	19 (1)
Chronic obstructive lung disease (missing n=3)	40 (0.7)	25 (0.7)	15 (0.8)
Autoimmune disease (missing n=332)	38 (0.7)	31 (0.8)	7 (0.4)
Symptoms (missing n=4), n (%)			
Sputum	1610 (28.8)	1114 (29.9)	496 (26.6)
Fever	1300 (23.2)	852 (22.8)	448 (24)
Dyspnea	662 (11.8)	454 (12.2)	208 (11.2)
Diarrhea	516 (9.2)	345 (9.3)	171 (9.2)
Nausea or vomiting	244 (4.4)	168 (4.5)	76 (4.1)
Fatigue	233 (4.2)	164 (4.4)	69 (3.7)
Mental disturbance, n (%)	32 (0.6)	22 (0.6)	10 (0.5)
Systolic blood pressure (mmHg) (missing n=135), n (%)			
<120	1306 (23.3)	907 (24.3)	399 (21.4)
120-129	1138 (20.3)	733 (19.7)	405 (21.7)
130-139	1084 (19.4)	705 (18.9)	379 (20.3)
140-159	1418 (25.3)	960 (25.7)	458 (24.6)
≥160	513 (9.2)	330 (8.8)	183 (9.8)
Diastolic blood pressure (mmHg) (missing n=135), n (%)			
<80	2102 (37.6)	1401 (37.6)	701 (37.6)

Variable	Total cohort (N=5594), n (%)	Subcohorts	
		Development cohort (n=3729), n (%)	Validation cohort (n=1865), n (%)
80-89	1797 (32.1)	1201 (32.2)	596 (32)
90-99	1056 (18.9)	686 (18.4)	370 (19.8)
≥100	504 (9)	347 (9.3)	157 (8.4)
Heart rate (per min) (missing n=122), mean (SD)	85.8 (15.1)	85.8 (15.1)	85.8 (15.1)
<110, n (%)	5136 (91.8)	3,374 (90.5)	1709 (91.6)
≥110, n (%)	336 (6)	272 (7.3)	117 (6.3)
Body temperature (°C) (missing n=37), mean (SD)	36.9 (0.6)	36.9 (0.6)	36.9 (0.6)
<38°C, n (%)	5348 (95.6)	3523 (94.5)	1752 (93.9)
≥38°C, n (%)	209 (3.7)	179 (4.8)	103 (5.5)
Baseline hemogram			
Hemoglobin (g/dL) (missing n=1519), mean (SD)	13.3 (1.8)	13.3 (1.7)	13.3 (1.8)
≥12.5, n (%)	2882 (51.5)	1923 (51.6)	959 (51.4)
<12.5, n (%)	1193 (21.3)	773 (20.7)	420 (22.5)
Absolute lymphocyte count (per mm^3) (missing n=1542), mean (SD)	1691 (1,054)	1697 (955)	1,681 (1,225)
≥1000, n (%)	3266 (58.4)	2161 (58)	1105 (59.2)
<1000, n (%)	786 (14.1)	518 (13.9)	268 (14.4)
Platelet count (per mm^3) (missing n=1517), mean (SD)	236,814 (82,846)	238,377 (82,789)	233,760 (82,900)
≥100,000, n (%)	3986 (71.3)	2634 (70.6)	1352 (72.5)
<100,000, n (%)	91 (1.6)	62 (1.7)	29 (1.6)
Follow-up (days), median (95% CI)	25 (24-25)	25 (24-25)	25 (24-25)

[a]Cases that achieved complete cure of cancer were excluded.

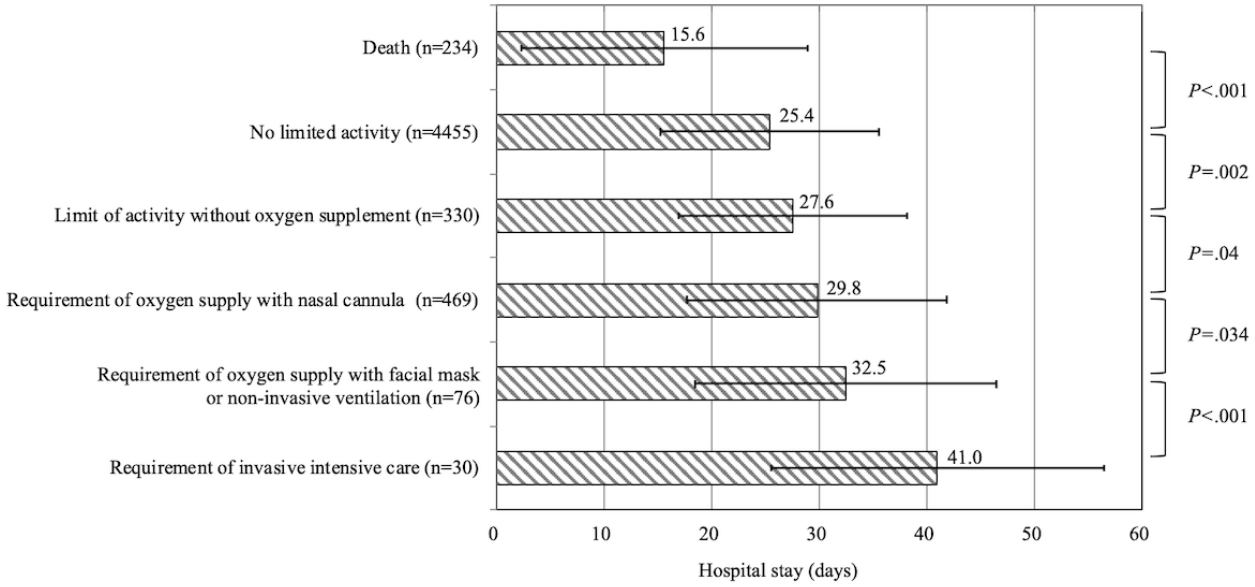
[b]Cases with chronic hepatitis were included in this category.

Clinical Course, Outcome, and Length of Hospital Stay in the Total Cohort

Among the 5594 patients included in the analysis, 234 (4.2%) died after a COVID-19 diagnosis was made, resulting in a cohort case fatality rate of 4.2%. Excluding death, the maximal clinical disease severity during hospitalization was as follows: (1) no limit of activity in 79.6% (4455/5594) of the patients, (2) limited activity without oxygen supplementation in 5.9% (330/5594) of the patients, (3) requirement of oxygen supply with nasal cannula in 8.4% (469/5594) of the patients, (4) requirement of oxygen supply with facial mask or advanced device (such as noninvasive ventilation or high flow oxygen therapy) in 1.4% (76/5594) of the patients, and (5) requirement of invasive intensive care such as invasive ventilation for acute respiratory distress syndrome or ECMO for multiple organ failure in 0.5% (30/5594) of the patients (Figure 2).

Overall, the mean duration of hospital stay was 25.6 (SD 11.0) days. Hospital stay was significantly longer among survivors (mean 26.1, SD 10.7 days) than among nonsurvivors (mean 15.6, SD 13.3 days; P<.001). As shown in Figure 2, the higher the severity of clinical course among survivors, the longer was their hospital stay: no limit of activity, mean 25.4 (SD 10.2) days; limited activity without oxygen supplementation, mean 27.6 (SD 10.6) days; oxygen supply with nasal cannula, mean 29.8 (SD 12.0) days; oxygen supply with facial mask or advanced device, mean 32.5 (SD 14.0) days; and invasive intensive care groups, mean 41.0 (SD 15.4) days.

Figure 2. Maximal COVID-19 disease severity and duration of hospital stay. Hospital stay duration differed significantly according to maximal disease severity; the mean length of hospital stay is indicated.



Analysis of Factors Associated With Overall Survival in the Development Cohort

The univariable analysis identified the following potential factors associated with poor overall survival: age (≥70 or 50-69 years vs <50 years); sex (male); comorbidities such as hypertension, diabetes, dementia, chronic cardiac disease, cancer in active treatment, chronic pulmonary disease, and chronic renal failure; dyspnea, fatigue, mental disturbance, high systolic blood pressure (≥140 mmHg), low diastolic blood pressure (<80 mmHg), tachycardia (heart rate ≥110/min), and fever (≥38°C) at the time of diagnosis; and cytopenia (hemoglobin level <12.5 g/dL, absolute lymphocyte count [ALC] <1000/mm^3, and platelet count <100,000/mm^3), as shown in Table 2.

Table 2. Univariable analysis for potential factors associated with overall survival in the development cohort. All values presented in the table represent data collected at the time of initial COVID-19 diagnosis.

Factor	Number of patients (n=3729), n (%)	Overall survival rate, % (95% CI)		*P* value
		At 14 days	At 28 days	
Age (years)				<.001
<50	1782 (47.8)	99.9 (99.8-100)	99.8 (99.5-100)	
50-69	1373 (36.8)	98.6 (98-99.2)	98 (97.2-98.8)	
≥70	574 (15.4)	87 (84.3-89.8)	81.6 (78.3-85)	
Sex				.01
Female	2193 (58.8)	97.8 (97.2-98.4)	96.8 (96-97.6)	
Male	1536 (41.2)	96.9 (96-97.8)	95.2 (94.1-96.4)	
Comorbidity				
Hypertension				<.001
No	2932 (78.6)	98.7 (98.3-99.1)	98 (97.4-98.6)	
Yes	795 (21.3)	92.7 (90.9-94.6)	89.7 (87.4-91.9)	
Diabetes				<.001
No	3275 (87.8)	98.3 (97.8-98.7)	97.3 (96.7-98)	
Yes	452 (12.1)	91.5 (88.9-94.1)	88.1 (85-91.3)	
Dementia				<.001
No	3359 (90.1)	98.3 (97.9-98.7)	97.3 (96.7-97.9)	
Yes	150 (4)	74.5 (67.8-81.8)	67.4 (60.1-75.6)	
Chronic cardiac disease [a]				<.001
No	3582 (96.1)	97.7 (97.2-98.2)	96.7 (96-97.3)	
Yes	147 (3.9)	91.1 (86.6-95.8)	85.8 (79.9-92)	
Cancer in active treatment				.03
No	3636 (97.5)	97.5 (97-98)	96.3 (95.7-97)	
Yes	90 (2.4)	95.5 (91.2-99.9)	89.8 (82.5-97.7)	
Chronic pulmonary disease [b]				<.001
No	3628 (97.3)	97.6 (97.1-98.1)	96.5 (95.9-97.2)	
Yes	101 (2.7)	92 (86.8-97.5)	85.3 (77.9-93.3)	
Chronic hepatic disease				.6
No	3463 (92.9)	97.3 (96.7-97.8)	96 (95.2-96.7)	
Yes	47 (1.3)	97.9 (93.8-100)	95.3 (89.1-100)	
Chronic renal failure				<.001
No	3691 (99)	97.5 (97-98)	96.3 (95.7-97)	
Yes	36 (1)	85.7 (74.9-98.1)	81.2 (68.4-96.4)	
Autoimmune disease				.56
No	3476 (93.2)	97.3 (96.7-97.8)	96 (95.3-96.7)	
Yes	31 (0.8)	96.8 (90.8-100)	90.7 (78.7-100)	
Symptoms				
Sputum				.59
No	2612 (70)	97.4 (96.8-98.1)	96 (95.2-96.9)	
Yes	1114 (29.9)	97.4 (96.4-98.3)	96.5 (95.4-97.7)	
Dyspnea				<.001

Factor	Number of patients (n=3729), n (%)	Overall survival rate, % (95% CI)		*P* value
		At 14 days	At 28 days	
No	3272 (87.7)	98.3 (97.9-98.8)	97.4 (96.8-98)	
Yes	454 (12.2)	90.9 (88.2-93.6)	87.6 (84.4-90.9)	
Diarrhea				.81
No	3381 (90.7)	97.4 (96.8-97.9)	96.1 (95.4-96.8)	
Yes	345 (9.3)	98 (96.5-99.5)	96.8 (94.8-98.8)	
Nausea/vomiting				.2
No	3558 (95.4)	97.5 (97-98.1)	96.3 (95.6-97)	
Yes	168 (4.5)	95.1 (91.8-98.5)	94.3 (90.8-98)	
Fatigue				.006
No	3562 (95.5)	97.5 (97-98.1)	96.4 (95.7-97.1)	
Yes	164 (4.4)	95 (91.7-98.5)	91.9 (87.6-96.5)	
Mental disturbance				<.001
No	3704 (99.3)	97.7 (97.3-98.2)	96.5 (95.9-97.2)	
Yes	22 (0.6)	45.5 (28.8-71.8)	40.4 (24.2-67.5)	
Systolic blood pressure (mmHg)				.02
<140	2345 (62.9)	97.8 (97.2-98.4)	96.8 (96-97.6)	
≥140	1290 (34.6)	96.9 (95.9-97.9)	95.3 (94-96.6)	
Diastolic blood pressure (mmHg)				.01
<80	1401 (37.6)	96.7 (95.7-97.6)	94.9 (93.7-96.2)	
≥80	2234 (59.9)	98 (97.4-98.6)	97.1 (96.3-97.8)	
Heart rate (per min)				.005
<110	3374 (90.5)	97.7 (97.2-98.2)	96.4 (95.7-97.1)	
≥110	272 (7.3)	94.4 (91.6-97.2)	93.5 (90.6-96.6)	
Body temperature (°C)				<.001
<38	3523 (94.5)	97.8 (97.3-98.3)	96.4 (95.8-97.1)	
≥38	179 (4.8)	93.2 (89.6-97)	93.2 (89.6-97)	
Baseline hemogram				
Hemoglobin (g/dL)				<.001
≥12.5	1923 (51.6)	97.9 (97.2-98.5)	96.8 (95.9-97.7)	
<12.5	773 (20.7)	93.6 (91.8-95.3)	90.8 (88.7-93.1)	
Absolute lymphocyte count (per mm ^{3})				<.001
≥1000	2161 (58)	98.3 (97.8-98.9)	97.7 (97-98.4)	
<1000	518 (13.9)	89.7 (87.1-92.4)	84.3 (81-87.8)	
Platelet count (per mm ^{3})				<.001
≥100,000	2634 (70.6)	96.9 (96.2-97.5)	95.2 (94.3-96.1)	
<100,000	62 (1.7)	85.3 (76.9-94.7)	83.5 (74.7-93.4)	

[a]Chronic cardiac disease was a composite variable including heart failure and cardiac disease.

[b]Chronic pulmonary disease was a composite variable including asthma and chronic obstructive lung disease.

COVID-19 Prognosis Score for Predicting Overall Survival

In the bootstrap analysis, we identified that older age (50-69 or ≥70 years) and comorbidities, including dementia, chronic renal failure, presentation of dyspnea, mental disturbance at diagnosis, and ALC <1,000 /mm^3 were significantly associated with poor overall survival. Assigned risk scores obtained by rounding the log_e scale of hazard ratio are shown in Table 3: age (50-69 years, 2 points; ≥70 years, 3 points), underlying dementia (1 point), chronic renal failure (1 point), dyspnea (1 point), mental disturbance (1 point), ALC <1000/mm^3 (1 point). We determined the COVID-19 prognosis score (COPS) based on the risk scores obtained for each patient and summing the respective scores of the 6 parameters. The total COPS ranged between 0 and 8.

We explored the clinical prediction score in the validation cohort using the AUROC curve analysis, which resulted in an AUROC value of 0.918 (95% CI 0.91-0.927) for the 14-day overall survival rate and 0.896 (95% CI 0.872-0.911) for the 28-day overall survival rate, indicating a reliable discrimination through the COPS system (Figure 3). Thereafter, we applied the scoring system to the total cohort, which resulted in a score range of 0-7 points (Figure 4A). This scoring system discriminately classified the patients into 8 groups. The 28-day overall survival rates were predicted as 99.9% (95% CI 99.7-100) in the 0-point group (n=2348), 99.7% (95% CI 99.1-100) in the 1-point group (n=317), 99.6% (95% CI 92.9-99.9) in the 2-point group (n=1511), 95.4% (95% CI 93.9-97.1) in the 3-point group (n=815), 82.3% (95% CI 78.5-86.4) in the 4-point group (n=395), 60% (95% CI 39.2-52.8) in the 5-point group (n=170), 32.9% (95% CI 20.6-52.7) in the 6-point group (n=36), and 50% (95% CI 12.5-100) in the 7-point group (n=2) (P<.001; Figure 4A).

We then determined the risk groups based on the final COPS system; these included low-risk (0-2 points, n=4167), intermediate-risk (3 points, n=774), high-risk (4 points, n=321), and very-high risk (≥5 points, n=98) groups. The 28-day overall survival rates for these groups were as follows: low-risk group, 99.8% (95% CI 99.6-99.9); intermediate-risk group, 95.4% (95% CI 93.9-97.1); high-risk group, 82.3% (95% CI 78.5-86.4); and very-high risk group, 55.1% (95% CI 48.5-62.5) (P<.001; Figure 4B). The developed COPS calculator is available online [17].

Furthermore, a significant increase in the length of hospital stay was observed as the risk group advanced: low-risk group, mean 25.4 (SD 10.4) days; intermediate-risk group, mean 27.2 (SD 10.9) days; and high-risk or very high-risk groups, mean 30.8 (SD 11.9) days (P<.001).

Table 3. The final scoring model in the development cohort. All values presented in the table represent data collected at the time of initial COVID-19 diagnosis.

Factor	Adjusted hazard ratio (95% CI)[a]	*P* value[a]	Log_e value of hazard ratio	Final score
Age (years)				
<50	1 (reference)	N/A[b]	0	0
50-69	6.7 (1.09-43.93)	.047	1.831	2
≥70	26.03 (4.26-169.8)	<.001	3.186	3
Sex				
Female	1 (reference)	N/A	N/A	N/A
Male	1.35 (0.85-2.15)	.3	N/A	N/A
Comorbidity				
Hypertension				
No	1 (reference)	N/A	N/A	N/A
Yes	1.21 (.75-1.94)	.48	N/A	N/A
Diabetes				
No	1 (reference)	N/A	N/A	N/A
Yes	1.71 (1.08-2.7)	.61	N/A	N/A
Dementia				
No	1 (reference)	N/A	0	0
Yes	3.92 (2.33-6.61)	<.001	1.35	1
Chronic cardiac disease[c]				
No	1 (reference)	N/A	N/A	N/A
Yes	1.15 (.61-2.15)	.57	N/A	N/A
Cancer in active treatment				
No	1 (reference)	N/A	N/A	N/A
Yes	1.83 (.72-4.71)	.30	N/A	N/A
Chronic pulmonary disease[d]				
No	1 (reference)	N/A	N/A	N/A
Yes	1.76 (.857-3.625)	.25	N/A	N/A
Chronic renal failure				
No	1 (reference)	N/A	0	0
Yes	3.48 (1.39-8.85)	.045	1.205	1
Symptoms				
Dyspnea				
No	1 (reference)	N/A	0	0
Yes	2.31 (1.44-3.69)	.006	.821	1
Fatigue				
No	1 (reference)	N/A	N/A	N/A
Yes	.69 (.3-1.57)	.38	N/A	N/A
Mental disturbance				
No	1 (reference)	N/A	0	0
Yes	4.04 (1.72-9.53)	.04	1.268	1
Systolic blood pressure (mmHg)				

Factor	Adjusted hazard ratio (95% CI)[a]	*P* value[a]	Log_e value of hazard ratio	Final score
<140	1 (reference)	N/A	N/A	N/A
≥140	.93 (.57-1.54)	.59	N/A	N/A
Diastolic blood pressure (mmHg)				
<80	1 (reference)	N/A	N/A	N/A
≥80	1.03 (.63-1.68)	.63	N/A	N/A
Heart rate (per min)				
<110	1 (reference)	N/A	N/A	N/A
≥110	1.88 (.96-3.68)	.18	N/A	N/A
Body temperature (°C)				
<38	1 (reference)	N/A	N/A	N/A
≥38	.79 (.39-1.63)	.46	N/A	N/A
Baseline hemogram				
Hemoglobin (g/dL)				
≥12.5	1 (reference)	N/A	N/A	N/A
<12.5	1.14 (.71-1.83)	.55	N/A	N/A
Absolute lymphocyte count (per mm^3)				
≥1000	1 (reference)	N/A	0	0
<1000	2.71 (1.66-4.43)	<.001	.982	1
Platelet count (per mm^3)				
≥100,000	1 (reference)	N/A	N/A	N/A
<100,000	1.68 (.77-3.7)	.34	N/A	N/A

[a]Based on 2000 bootstrap samples.

[b]Not applicable.

[c]Chronic cardiac disease was a composite variable including heart failure and cardiac disease.

[d]Chronic pulmonary disease was a composite variable including asthma and chronic obstructive lung disease.

Figure 3. Receiver operating characteristics curve analysis of the newly developed COVID-19 prognosis score system in the validation cohort. A receiver operating characteristic curve analysis achieved an area under the curve value of (A) 0.918 (95% CI, 0.91-0.927) for 14-day survival and (B) 0.896 (95% CI 0.872-0.911) for 28-day survival.

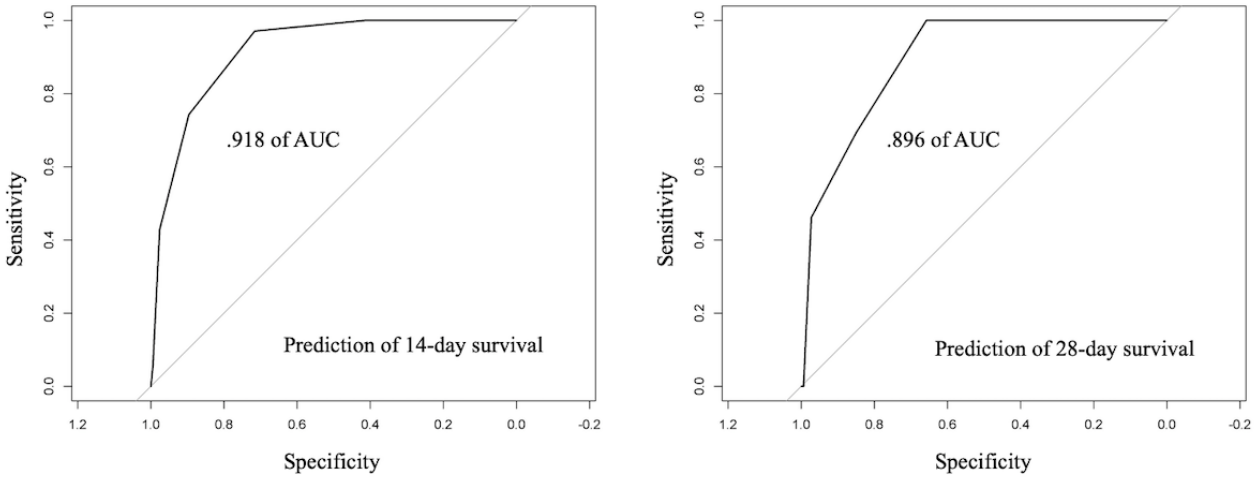
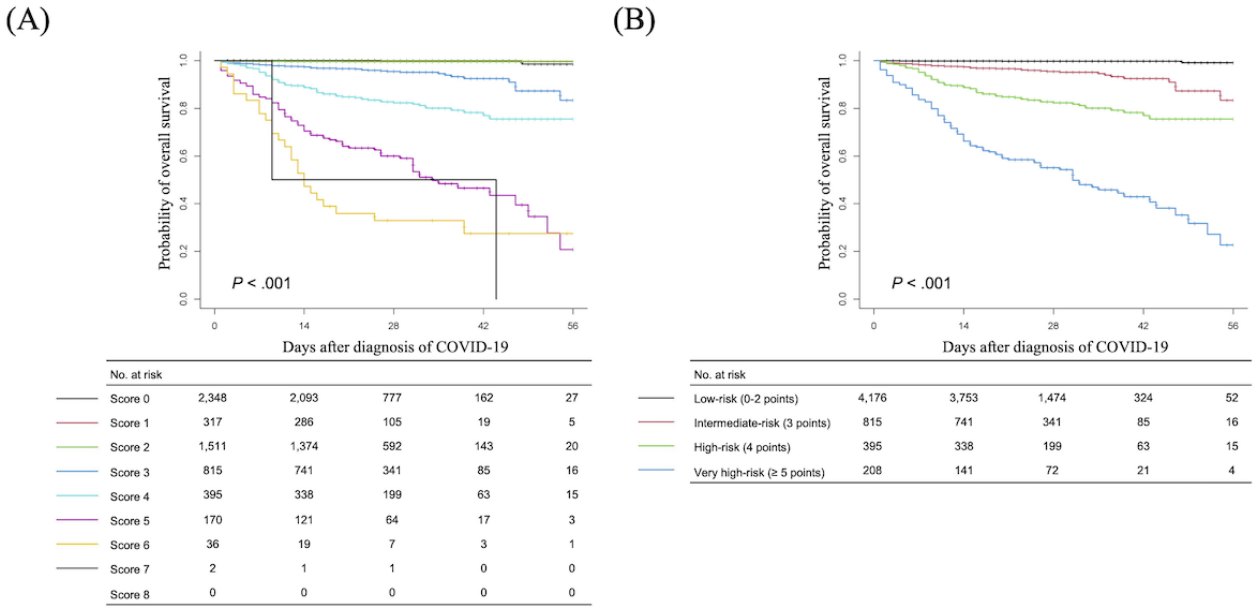


Figure 4. Probability of overall survival in patients with COVID-19 according to (A) each score and (B) the final COVID-19 prognosis score system.



Discussion

In this study, we developed a new scoring system (COPS) to predict the mortality of patients with COVID-19 by using nationwide data of 5594 patients with COVID-19. The COPS system comprises basic demographics, initial symptoms, vital signs, and hemogram results at diagnosis. The risk score was stratified into four risk groups: low-risk, intermediate-risk, high-risk, and very high-risk groups associated with 28-day overall survival rate probabilities of 99.8%, 95.4%, 82.3%, and 55.1%, respectively. The AUROC curve analysis indicated that the prediction ability of the COPS system was excellent in the validation cohort.

Among the comorbidities identified in this study, dementia had the highest impact on mortality (adjusted hazard ratio 3.92) followed by chronic renal failure (adjusted hazard ratio 3.48). Moreover, the prognosis was poor when mental disturbance (adjusted hazard ratio 4.04) was noted at the time of diagnosis or when the patient had underlying dementia (adjusted hazard ratio 3.92). These two factors are related to the presence of SARS-CoV-2 infection clusters in nursing homes or long-term care facilities between February and April 2020 in Korea. Therefore, there is a need to establish screening and infection control strategies for long-term health care facilities [18-21].

In this study, an ALC of <1000/mm^3 was found to affect the survival rate of patients with COVID-19. ALC has been used as a prognostic factor for common respiratory viruses, including respiratory syncytial virus or other viral reactivation in immunocompromised hosts [22]. A recent meta-analysis reported that lymphopenia on admission was associated with poor outcomes in patients with COVID-19 [23]. Further immunological studies on patients with COVID-19 are needed to elucidate the mechanism of lymphopenia and T cell reactivation, as well as cytokines [24].

Regarding disease severity, approximately 87% of patients did not need oxygen supplementation. Approximately 10% of the patients required oxygen supplementation; of these, 25% received oxygen via a simple mask or mechanical ventilators. From the perspective of a national strategy for new infectious disease crisis, it is important to determine the proportion of critically ill patients and the length of hospital stay according to disease severity in order to prepare medical resources, such as critical care beds. This study found that the hospitalization period was significantly longer among survivors than among nonsurvivors. Moreover, among the survivors, the length of hospital stay was directly associated with disease severity.

In this cohort, the infection fatality rate after COVID-19 diagnosis was 4.18% (234/5594). However, until April 30, 2020, the cumulative number of confirmed COVID-19 cases in Korea was 10,765 with 247 deaths, representing an actual mortality rate of 2.29% during the same period [25]. As of November 6, 2020, the total cumulative number of confirmed COVID-19 cases and deaths were 27,195 and 476, respectively, representing an infection fatality rate of 1.75% in Korea [25]. This disparity in mortality rates can be attributed to the fact that not all data were reported at the time this cohort was released. However, considering the large number of patients included in this cohort and collection of most deceased cases, the findings of this study are still meaningful and carry little statistical bias. In addition, remdesivir was not available in Korea during the study period, and infectious disease prevention and control measures were less established in the early phase of the COVID-19 pandemic. Thus, several mass infection episodes might have caused the relatively high mortality rate during the early phase of the pandemic in Korea.

Several studies have attempted to determine predictive factors for severe or fatal COVID-19 cases. A study on risk factors for fatal COVID-19 cases performed in China proposed a scoring system comprising age, procalcitonin, aspartate aminotransferase level, coronary heart disease, and cerebrovascular disease; this system was developed using data of 1590 inpatients with COVID-19 collected until January 2020. The nomogram showed discriminatory power with a C-index of 0.91 to predict survival

[13]. Another study analyzed the risk of intensive care unit admissions and deaths among 4997 individuals under investigation at an academic hospital in New York. The study used age, heart rate, procalcitonin, lactate dehydrogenase, heart failure, and chronic obstructive pulmonary disease to predict care in intensive care units and deaths, yielding an accuracy of 0.74 and 0.83, respectively [26]. In the United Kingdom, a similar study was conducted on 17 million individuals included in the OpenSAFELY database, a near–real-time primary care patient record, which pseudonymously identified 10,926 COVID-19–related deaths. The study found male sex, older age and deprivation, diabetes, and severe asthma as significant risk factors for death [27]. However, the study did not analyze survival and death among patients with a confirmed COVID-19 diagnosis. In addition to studies attempting to predict death in patients with COVID-19, other studies have focused on severity index to predict severe or critical cases [10,28-31]. More recently, a machine learning–based warning system for mortality risk prediction of patients with COVID-19 was reported, and timely risk stratification using multiple laboratory and clinical factors was improved [32].

Compared with the abovementioned studies, our study has a number of strengths. First, the risk factors for death were analyzed using a nationwide cohort comprising a large number of patients with COVID-19, which resulted in a scoring system that can be widely used for triaging laboratory-confirmed COVID-19 cases. Second, the COPS system was developed using easily accessible information, such as age, underlying disease, dyspnea, mental disturbance, and hemogram parameters. We believe that if a scoring system that uses only simple laboratory tests (eg, hemogram parameters) can be developed and still show good predictability, it will prove to be more cost-effective than systems including other biomarkers such as procalcitonin or cytokine levels. Third, the discriminatory power of our system for predicting death probability was excellent. Finally, this study further analyzed the length of hospital stay according to disease severity, which may assist in preparing medical resources based on patient classification. However, this study is limited by the lack of external verification for our scoring model. Therefore, the current model of overall survival for the diagnosis of COVID-19 would need to be validated in a future cohort.

In conclusion, our study provides a simple scoring system based on information collected at diagnosis for predicting mortality among patients with COVID-19 in a timely manner. Early triaging of patients with COVID-19 using the COPS system can provide new insights for risk-adaptive strategies and optimize the use of medical resources.

Acknowledgments

We would like to thank Si-Hyun Park and Seunghoon Han for their help with analyzing the data at Korea Disease Control and Prevention Agency. We acknowledge all the health care workers involved in the diagnosis and treatment of patients with COVID-19 in Korea. We thank Korea Disease Control & Prevention Agency, National Medical Center, and the Health Information Manager in hospitals for their effort in collecting the medical records. This work is funded by the Life Respecting Fund of Seoul St. Mary's hospital and Research Fund of Seoul St. Mary's Hospital, The Catholic University of Korea.

Authors' Contributions

SYC and SSP wrote the manuscript. SSP and MKS collected and analyzed the data and analyzed it. Drafted manuscript was reviewed by YYB, DGL and DWK. DGL and DWK designed the study, conceived the idea, and planned the project. SYC and SSP contributed equally to this work as first authors. DGL and DWK contributed equally to this work as corresponding authors. All authors approved the final version.

Conflicts of Interest

None declared.

References

1. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020 Mar;579(7798):270-273 [FREE Full text] [doi: 10.1038/s41586-020-2012-7] [Medline: 32015507]
2. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb 15;395(10223):470-473 [FREE Full text] [doi: 10.1016/S0140-6736(20)30185-9] [Medline: 31986257]
3. WHO Coronavirus Disease (COVID-19) Dashboard. World Health Organization. 2020. URL: https://covid19.who.int/ [accessed 2020-11-10]
4. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University & Medicine - Coronavirus Resource Center. 2020. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-11-10]
5. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]

6. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic transmission, the Achilles' heel of current strategies to control Covid-19. N Engl J Med 2020 May 28;382(22):2158-2160 [FREE Full text] [doi: 10.1056/NEJMe2009758] [Medline: 32329972]
7. Long Q, Tang X, Shi Q, Li Q, Deng H, Yuan J, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nat Med 2020 Aug;26(8):1200-1204. [doi: 10.1038/s41591-020-0965-6] [Medline: 32555424]
8. Cummings MJ, Baldwin MR, Abrams D, Jacobson SD, Meyer BJ, Balough EM, et al. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. The Lancet 2020 Jun;395(10239):1763-1770. [doi: 10.1016/s0140-6736(20)31189-2]
9. CDC COVID-19 Response Team. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) - United States, February 12-March 16, 2020. MMWR Morb Mortal Wkly Rep 2020 Mar 27;69(12):343-346 [FREE Full text] [doi: 10.15585/mmwr.mm6912e2] [Medline: 32214079]
10. Liu D, Cui P, Zeng S, Wang S, Feng X, Xu S, et al. Risk factors for developing into critical COVID-19 patients in Wuhan, China: A multicenter, retrospective, cohort study. EClinicalMedicine 2020 Aug;25:100471 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100471] [Medline: 32840491]
11. People with Certain Medical Conditions. Centers for Disease Control and Prevention - COVID-19. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html [accessed 2021-02-08]
12. Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. J Infect 2020 Aug;81(2):e16-e25 [FREE Full text] [doi: 10.1016/j.jinf.2020.04.021] [Medline: 32335169]
13. Chen R, Liang W, Jiang M, Guan W, Zhan C, Wang T, Medical Treatment Expert Group for COVID-19. Risk factors of fatal outcome in hospitalized subjects with coronavirus disease 2019 from a nationwide analysis in China. Chest 2020 Jul;158(1):97-105 [FREE Full text] [doi: 10.1016/j.chest.2020.04.010] [Medline: 32304772]
14. WHO COVID-19 Case definition. World Health Organization. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Surveillance_Case_Definition-2020.1 [accessed 2020-11-10]
15. Case definition and people subject to testing (As of June 25, 2020). Coronavirus Disease-19, Republic of Korea. URL: http://ncov.mohw.go.kr/en/baroView.do?brdId=11&brdGubun=112&dataGubun=&ncvContSeq=&contSeq=&board_id=&gubunn [accessed 2020-11-10]
16. Kim ES, Chin BS, Kang CK, Kim NJ, Kang YM, Choi J, Korea National Committee for Clinical Management of COVID-19. Clinical course and outcomes of patients with severe acute respiratory syndrome coronavirus 2 infection: a preliminary report of the first 28 patients from the Korean cohort study on COVID-19. J Korean Med Sci 2020 Apr 06;35(13):e142 [FREE Full text] [doi: 10.3346/jkms.2020.35.e142] [Medline: 32242348]
17. Prognosis score system to predict COVID-19 survival. Young Medical Research - COVID-19 Prognosis Scoring system (COPS). URL: https://ymdtech.kr/covid19-os [accessed 2021-02-08]
18. Nursing Homes and Long-Term Care Facilities. Centers for Disease Control and Prevention - COVID-19. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/nursing-home-long-term-care.html [accessed 2020-11-15]
19. Surveillance of COVID-19 at long-term care facilities in the EU/EEA. European Center for Disease Prevention and Control. 2020 May 19. URL: https://www.ecdc.europa.eu/en/publications-data/surveillance-COVID-19-long-term-care-facilities-EU-EEA [accessed 2021-02-08]
20. McMichael TM, Currie DW, Clark S, Pogosjans S, Kay M, Schwartz NG, Public Health–Seattle and King County, EvergreenHealth, and CDC COVID-19 Investigation Team. Epidemiology of Covid-19 in a long-term care facility in King County, Washington. N Engl J Med 2020 May 21;382(21):2005-2011 [FREE Full text] [doi: 10.1056/NEJMoa2005412] [Medline: 32220208]
21. Kim T, Choi MJ, Kim SB, Kim JY, Lee J, Oh HS, et al. Strategical preparedness and response actions in the healthcare system against coronavirus disease 2019 according to transmission scenario in Korea. Infect Chemother 2020 Sep;52(3):389-395 [FREE Full text] [doi: 10.3947/ic.2020.52.3.389] [Medline: 32757499]
22. Pochon C, Voigt S. Respiratory virus infections in hematopoietic cell transplant recipients. Front Microbiol 2018;9:3294 [FREE Full text] [doi: 10.3389/fmicb.2018.03294] [Medline: 30687278]
23. Huang I, Pranata R. Lymphopenia in severe coronavirus disease-2019 (COVID-19): systematic review and meta-analysis. J Intensive Care 2020;8:36 [FREE Full text] [doi: 10.1186/s40560-020-00453-4] [Medline: 32483488]
24. Chen Z, John Wherry E. T cell responses in patients with COVID-19. Nat Rev Immunol 2020 Sep;20(9):529-536 [FREE Full text] [doi: 10.1038/s41577-020-0402-6] [Medline: 32728222]
25. Coronavirus Disease-19 cases in Korea. Korea Disease Control and Prevention Agency. URL: http://ncov.mohw.go.kr/en/ [accessed 2020-11-10]
26. Zhao Z, Chen A, Hou W, Graham JM, Li H, Richman PS, et al. Prediction model and risk scores of ICU admission and mortality in COVID-19. PLoS One 2020;15(7):e0236618 [FREE Full text] [doi: 10.1371/journal.pone.0236618] [Medline: 32730358]
27. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020 Aug;584(7821):430-436. [doi: 10.1038/s41586-020-2521-4] [Medline: 32640463]
28. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, China Medical Treatment Expert Group for COVID-19. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19.

JAMA Intern Med 2020 Aug 01;180(8):1081-1089 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2033] [Medline: 32396163]
29. Bhargava A, Fukushima EA, Levine M, Zhao W, Tanveer F, Szpunar SM, et al. Predictors for severe COVID-19 infection. Clin Infect Dis 2020 Nov 05;71(8):1962-1968 [FREE Full text] [doi: 10.1093/cid/ciaa674] [Medline: 32472676]
30. Haimovich AD, Ravindra NG, Stoytchev S, Young HP, Wilson FP, van Dijk D, et al. Development and validation of the quick COVID-19 severity index: a prognostic tool for early clinical decompensation. Ann Emerg Med 2020 Oct;76(4):442-453 [FREE Full text] [doi: 10.1016/j.annemergmed.2020.07.022] [Medline: 33012378]
31. Altschul DJ, Unda SR, Benton J, de la Garza Ramos R, Cezayirli P, Mehler M, et al. A novel severity score to predict inpatient mortality in COVID-19 patients. Sci Rep 2020 Oct 07;10(1):16726 [FREE Full text] [doi: 10.1038/s41598-020-73962-9] [Medline: 33028914]
32. Gao Y, Cai G, Fang W, Li H, Wang S, Chen L, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. Nat Commun 2020 Oct 06;11(1):5033 [FREE Full text] [doi: 10.1038/s41467-020-18684-2] [Medline: 33024092]

Abbreviations

ALC: absolute lymphocyte count
AUROC: area under the receiver operating characteristic curve
COPS: COVID-19 prognosis score
ECMO: extracorporeal membrane oxygenation
KDCA: Korea Disease Control and Prevention Agency

Edited by C Basch; submitted 04.12.20; peer-reviewed by CY Lin, J Liu; comments to author 29.12.20; revised version received 19.01.21; accepted 03.02.21; published 22.02.21.

Please cite as:
Cho SY, Park SS, Song MK, Bae YY, Lee DG, Kim DW
Prognosis Score System to Predict Survival for COVID-19 Cases: a Korean Nationwide Cohort Study
J Med Internet Res 2021;23(2):e26257
URL: https://www.jmir.org/2021/2/e26257
doi:10.2196/26257
PMID:33539312

Original Paper

Using Tweets to Understand How COVID-19–Related Health Beliefs Are Affected in the Age of Social Media: Twitter Data Analysis Study

Hanyin Wang[1], BMed; Yikuan Li[1], BSc, MSc; Meghan Hutch[1], BSc; Andrew Naidech[2], MD, MSPH; Yuan Luo[1], PhD

[1]Department of Preventive Medicine, Northwestern University, Chicago, IL, United States

[2]Department of Neurology, Northwestern Universtity, Chicago, IL, United States

Corresponding Author:
Yuan Luo, PhD
Department of Preventive Medicine
Northwestern University
750 N Lakeshore Dr
Chicago, IL, 60611
United States
Phone: 1 312 503 5742
Email: yuan.luo@northwestern.edu

Abstract

Background: The emergence of SARS-CoV-2 (ie, COVID-19) has given rise to a global pandemic affecting 215 countries and over 40 million people as of October 2020. Meanwhile, we are also experiencing an infodemic induced by the overabundance of information, some accurate and some inaccurate, spreading rapidly across social media platforms. Social media has arguably shifted the information acquisition and dissemination of a considerably large population of internet users toward higher interactivities.

Objective: This study aimed to investigate COVID-19-related health beliefs on one of the mainstream social media platforms, Twitter, as well as potential impacting factors associated with fluctuations in health beliefs on social media.

Methods: We used COVID-19-related posts from the mainstream social media platform Twitter to monitor health beliefs. A total of 92,687,660 tweets corresponding to 8,967,986 unique users from January 6 to June 21, 2020, were retrieved. To quantify health beliefs, we employed the health belief model (HBM) with four core constructs: perceived susceptibility, perceived severity, perceived benefits, and perceived barriers. We utilized natural language processing and machine learning techniques to automate the process of judging the conformity of each tweet with each of the four HBM constructs. A total of 5000 tweets were manually annotated for training the machine learning architectures.

Results: The machine learning classifiers yielded areas under the receiver operating characteristic curves over 0.86 for the classification of all four HBM constructs. Our analyses revealed a basic reproduction number R_0 of 7.62 for trends in the number of Twitter users posting health belief–related content over the study period. The fluctuations in the number of health belief–related tweets could reflect dynamics in case and death statistics, systematic interventions, and public events. Specifically, we observed that scientific events, such as scientific publications, and nonscientific events, such as politicians' speeches, were comparable in their ability to influence health belief trends on social media through a Kruskal-Wallis test (P=.78 and P=.92 for perceived benefits and perceived barriers, respectively).

Conclusions: As an analogy of the classic epidemiology model where an infection is considered to be spreading in a population with an R_0 greater than 1, we found that the number of users tweeting about COVID-19 health beliefs was amplifying in an epidemic manner and could partially intensify the infodemic. It is "unhealthy" that both scientific and nonscientific events constitute no disparity in impacting the health belief trends on Twitter, since nonscientific events, such as politicians' speeches, might not be endorsed by substantial evidence and could sometimes be misleading.

(J Med Internet Res 2021;23(2):e26302) doi:10.2196/26302

KEYWORDS

COVID-19; social media; health belief; Twitter; infodemic; infodemiology; machine learning; natural language processing

Introduction

Beginning in December 2019, the outbreak of SARS-CoV-2 rapidly evolved into a global pandemic [1-3]. As of the writing of this paper, over 40 million cases and 1 million deaths from 215 countries or regions have been confirmed [4]. However, spreading faster than the virus is information. Sylvie Briand, Director of Infectious Hazards Management at the World Health Organization (WHO)'s Health Emergencies Programme, pointed out "We know that every outbreak will be accompanied by a kind of tsunami of information, but also within this information you always have misinformation, rumors, etc." [5]. The WHO used the term *infodemic* to describe the overabundance of information and misinformation occurring during the COVID-19 pandemic. Though the term infodemic was first coined in 2002 [6], the concerns over infodemics have recently become dramatic with the amplification effect from social media. The WHO held the first Infodemiology Conference in June 2020 as the phenomenon had escalated to a level that required a coordinated response [7]. Even though we cannot avoid an infodemic, we can still manage it. Previous studies and commentaries proposed several perspectives to detect and fight the COVID-19 infodemic [5,8-10]. However, one of the critical points absent from these studies is an investigation of health beliefs. Understanding how the general public's health beliefs are expressed and altered can facilitate our management of both the pandemic and infodemic. In conjunction, it is also essential to evaluate any concurrent or ongoing interventions.

The health belief model (HBM) quantifies health beliefs [11-13]. The HBM was developed to investigate people's beliefs about health problems. It consists of the following four core constructs that can be tailored for given hypotheses: (1) perceived susceptibility, (2) perceived severity, (3) perceived benefits, and (4) perceived barriers. The HBM has been widely used to investigate individual opinions toward diseases and interventional approaches, such as HIV risk behaviors [14], human papillomavirus (HPV) vaccines [15], and the gender difference in food choices [16]. In those cases, the HBM was employed to evaluate people's beliefs toward the given health problem and their perceived benefits or barriers of action, for which each of the core constructs of the HBM is assessed based on the corresponding definitions. During pandemics, researchers have employed the HBM to investigate the health beliefs toward public interventional policies, such as stay-at-home orders [17]; to analyze public health communication on Instagram during the Zika outbreak [18]; to examine public perceptions of physical distancing [19]; and to guide community pharmacists in their communication with patients [20]. However, because these are survey-based or merely commentary studies, results are limited to the analyzed population and, therefore, may be biased. In this study, we expanded and diversified our study population by using crowdsourcing data from one of the mainstream social media platforms, Twitter, in order to investigate the health beliefs of the general public toward COVID-19 and its potential treatments.

In addition to quantifying health beliefs, we aimed to identify factors influencing fluctuations in public opinions. For instance, the pandemic dynamics (ie, the number of cases and deaths due to COVID-19) constitute one of the leading factors influencing attitudes toward the pandemic. Additionally, interventional government policies may also impact the opinions of the general public. Furthermore, it is reasonable to believe that health belief–related posts can also be self-regulated as a consequence of their nature to induce or soothe panic for readers. Potential treatments trigger massive discussions as well, such as the debate over the appropriate use of the antimalarial drug hydroxychloroquine (HCQ) or chloroquine (CQ), advocated for by the then American President as a "game changer," which was then subsequently discarded. Public attitudes regarding potential treatments may be altered by public events, such as the news or politicians' speeches. Furthermore, rapidly emerging scientific publications can also influence the point of view of the general public. In this paper, we aim to identify factors that impact health beliefs on social media, which may serve as a probe for identifying better strategies to manage both the pandemic and the infodemic.

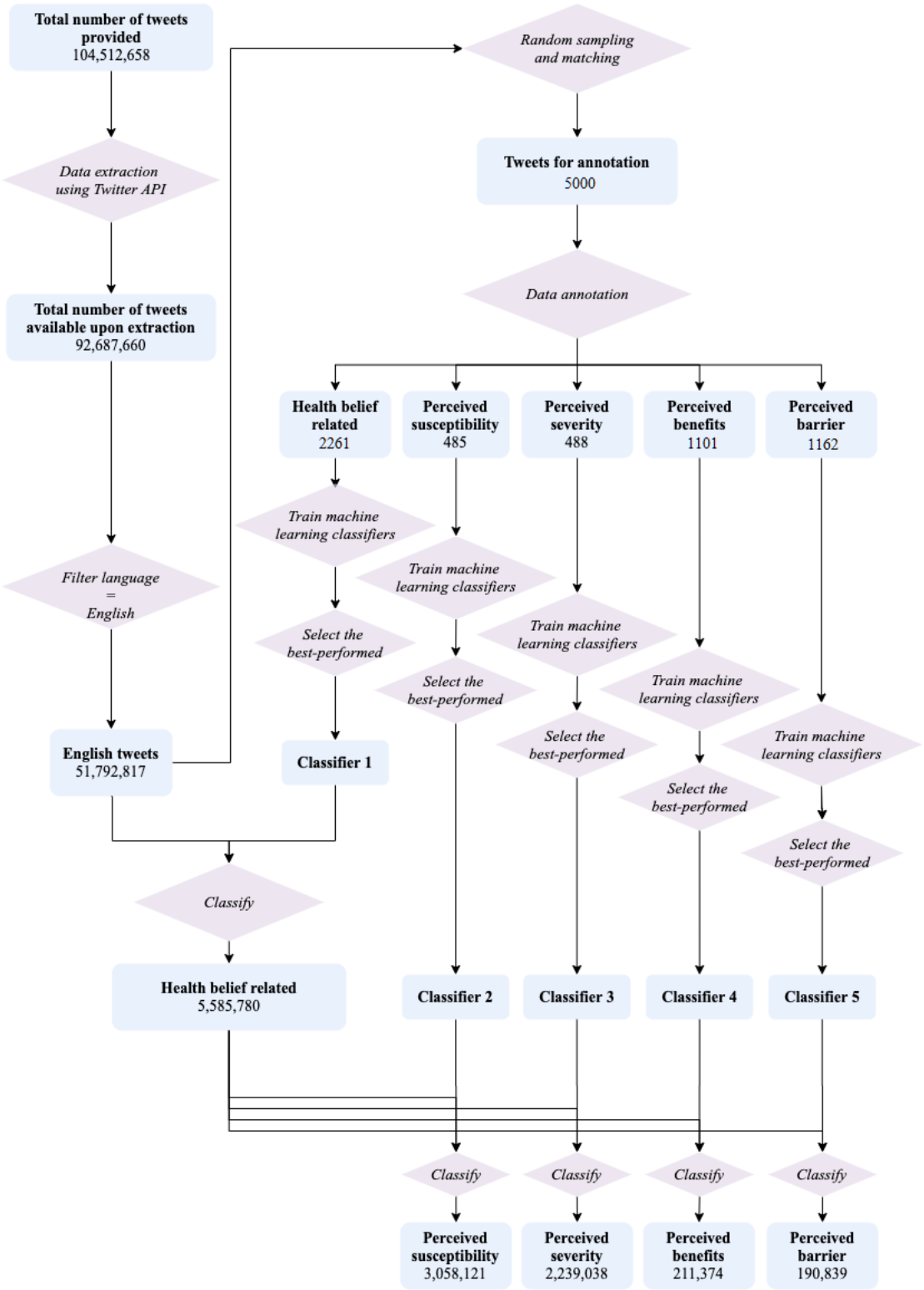
Contributions of this study include the following:

1. An evaluation of utilizing a mainstream social media platform, Twitter, to facilitate a comprehensive understanding of health beliefs toward COVID-19 and potential treatments.
2. A publicly available data set annotated by multiple professionals for studying the health beliefs related to COVID-19 and potential treatments.
3. Identification and comparison of factors that influence health beliefs toward COVID-19 and potential treatments, HCQ or CQ in particular.
4. An extendable framework for monitoring the general public's health beliefs during a pandemic and infodemic, which could be feasibly transferred to facilitate the management of future infodemic outbreaks, such as when COVID-19 vaccines become available to the public.

Methods

The entire workflow of data extraction, filtering, and classification is illustrated in Figure 1.

Figure 1. The workflow of data selection, data annotation, and machine learning classifier training and implementation. API: application programming interface.



Data

We used version 15.0 of the COVID-19 Twitter chatter data set constructed by the Panacea Lab [21], which collected all COVID-19-related tweets between January 6 and June 21, 2020. The provided data set only contains the identifiers of corresponding tweets, so we used the application programming interface (API) provided by Twitter to extract the full content of each tweet. The Social Media Mining Toolkit provided by the Panacea Lab was used to hydrate the data set [22]. There is no limitation regarding the days prior to the extraction. A language filter (ie, "lang" attribute in the tweet object = "en") was then applied to identify tweets written in the English language.

Data Annotation for Constructing the Health Belief Model

We employed the HBM to quantify health beliefs. As mentioned above, it consists of the following four core constructs that can be tailored for given hypotheses: (1) perceived susceptibility, (2) perceived severity, (3) perceived benefits, and (4) perceived barriers. The HBM was developed to investigate people's beliefs about health problems and has been widely used to investigate individual opinions toward diseases and interventional approaches, such as HIV risk behaviors [14], HPV vaccines [15], and the gender difference in food choices [16]. Specifically, for perceived benefits and barriers, we focused on HCQ or CQ, the antimalarial drug advocated for by the then American President as a "game changer," which was then subsequently discarded. Tweets were labeled as positive or negative for being related to HBM, meaning they could be mapped to at least one of the four aforementioned constructs. Thus, each tweet could potentially have up to five labels. The annotation process was performed by three senior PhD students in biomedical informatics (HW, YLi, and MH). All annotators classified the first 500 tweets individually, then reconciled different opinions and built final annotation rules. The definitions for each construct of the HBM are described in Table 1 with example tweets. Based on the rule, HW and YLi annotated the rest of 5000 tweets independently and evaluated the agreement using the Cohen κ score [23]. Finally, MH resolved the divergent annotations between HW and YLi with further consideration. We made the data set with the 5000 annotated tweets available for researchers [24]. To protect the privacy of Twitter users and per Twitter's policy, we did not include any tweet content in the data set. Instead, each tweet's unique identifier (ie, tweet ID) was provided.

Table 1. Health belief model (HBM) constructs, definitions, and example tweets.

Construct	Definitions of the construct	Example tweet
Perceived susceptibility	The assessment of the risk of getting COVID-19 infection	"Across the UK, 194,990 people had tested positive for coronavirus as of 9am on Tuesday, up from 190,584 at the same point on Monday. Find out how many cases there are in your area."
Perceived severity	The assessment of whether COVID-19 is a sufficient health concern	"US Recorded 1,297 Coronavirus Deaths in Past 24 Hours."
Perceived benefits	The benefits of HCQ[a] and CQ[b] in prevention or treatment of COVID-19	"Dr. Zelenko In NY has now treated 699 Coronavirus patients with 100% success using Hydroxychloroquine."
	Positive statements or reports about HCQ and CQ	
Perceived barriers	The side effects of HCQ and CQ	"Family of New York woman blames hydroxychloroquine combo for fatal heart attack."
	The unaffordable cost of HCQ and CQ	
	The inaccessibility of HCQ and CQ	
	Negative statements or reports about HCQ and CQ	
HBM related	Can be mapped to at least one of the four constructs above	Any of the above tweets are examples for this construct

[a]HCQ: hydroxychloroquine.

[b]CQ: chloroquine.

Machine Learning Classifiers

We trained machine learning classifiers on the annotated data, evaluated the performance, and automatically classified the 50 million tweets. The entire annotated data set was split into a training data set and a testing data set with a ratio of 8:2. Feature selection was applied by ignoring terms with a document frequency of less than 0.01 or greater than 0.99. Terms with only letters were considered, but terms were ignored when there were numbers or special characters. Before vectorization, we removed all the URLs; unified all the contractions, punctuation marks, and white spaces; and converted all terms to lowercase in the corpus. The free-text tweets were vectorized using both bag-of-words and term frequency–inverse document frequency (tf-idf) algorithms. A list of English stop words provided by the natural language toolkit [25] was used to rule out unrelated words. Next, 5-fold cross-validation was performed to select the best-suited classifier for the task. The machine learning classifiers that we experimented on included Ridge classifier; perceptron; passive-aggressive classifier; k-nearest neighbors classifier; random forest; support vector machine with linear kernel and l1 or l2 penalty; support vector machine with radial basis function, polynomial, or sigmoid kernel; stochastic gradient descent classifier with l1 or l2 or elastic net penalty; multinomial naïve Bayesian classifier; Bernoulli naïve Bayesian classifier; and logistic regression. Performance across classifiers was evaluated using the area under the receiver operating characteristic (AUROC) curve, and the classifier that yielded the highest AUROC was chosen (see Multimedia Appendix 1, Tables S1-S5). In total, five classifiers were built to construct the final HBM. First, a classifier was trained to classify whether a tweet was HBM related or not. Then, we collected all the tweets that were identified as HBM related for the following task. Lastly, we built four classifiers to label each core construct of the HBM separately.

The entire pipeline was built with Python 3.6.8 (Python Software Foundation). The bag-of-words algorithm, machine learning classifiers, and model evaluations were implemented with the scikit-learn, version 0.22.1, package [26].

The Overall Trend of Health Beliefs in Tweets

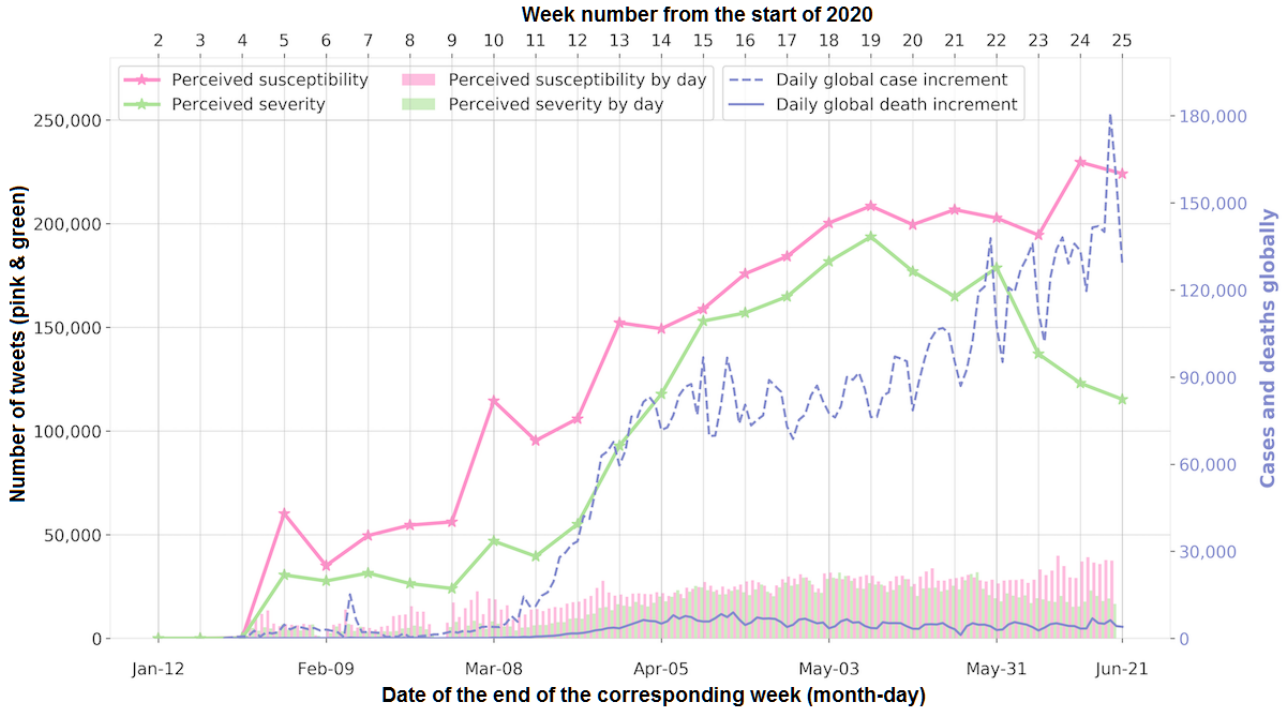
To quantify whether the information spread constituted an infodemic, we applied one of the classic measurements in

epidemiological models: the basic reproduction number, R_0. We employed the susceptible-infectious-recovered (SIR) model [27], for which the detailed calculation can be found in the Multimedia Appendix 1. In our case of an infodemic, we considered the users who tweeted about COVID-19 as the susceptible population; among this population, *being infected* meant a user tweeted about health beliefs defined in our HBM scope, and *recovering* then indicated that a user stopped tweeting about health beliefs. Thus, *contact with infected individuals* could be considered as reading health belief–related tweets posted by other users.

The Trend of Health Belief Toward the Disease

There are two core constructs in the HBM that focus specifically on the disease of interest: perceived susceptibility and perceived severity. We visualized these two constructs together with the dynamics of the pandemic in Figure 2. We observed a similar pattern in COVID-19 case dynamics and the number of tweets regarding perceived susceptibility, as well as in the dynamics of COVID-19 deaths and the number of tweets indicating perceived severity. For the first pair, we observed an earlier increase in the number of perceived susceptibility tweets prior to a surge in COVID-19 cases, while for the second pair, there was a delay in the increase of COVID-19 deaths compared with the number of perceived severity tweets. To investigate how many days the trend dynamics of health belief discussions followed or postponed the actual case or death increases, we calculated the Spearman correlation coefficient under various time lags (ie, for a 1-day lag, the correlation between the number of tweets and the COVID-19 situation was calculated by moving the COVID-19 trend 1 day forward). Moreover, we conducted a change point analysis using the dynamic programming algorithm [28] to detect the significant turning point of the trends.

Figure 2. Dynamics of perceived susceptibility and severity with COVID-19 case and death trends. The pink and green lines with "star" marks reflect the weekly cumulative number of tweets for perceived susceptibility and severity, while the pink and green bars on the x-axis indicate the daily number of tweets related to perceived susceptibility and severity. The global case and death dynamics have been available since January 22, 2020.



The Effect of Interventions

To evaluate the impact of interventions on the infodemic and the pandemic, we further investigated the lockdown in the United States. Because we were analyzing only tweets written in the English language, and there were systematic official lockdowns issued in the United States, we chose to study the effect of US-based interventions on health beliefs. The location information is not available for each tweet. We analyzed 136,641 tweets where the *place* was available—a variable in the tweet object that, when present, indicates that the tweet is associated with a place—and subsequently identified 54,164 tweets corresponding to the United States. We investigated the effect of interventions by visualizing the trends along with the timeline of lockdowns in the United States.

News in the Top Topics

To understand major topics in the tweets related to health beliefs, we extracted the top 10 phrases from the tweets each week. We considered unigrams and bigrams in this case. The frequency of each phrase was not only calculated as the count; instead, we used the tf-idf score to find the highlighted topics of each week. A higher tf-idf score is obtained if a given word or phrase frequently appears in one document but only appears in a small number of documents.

The Influence of Scientific and Nonscientific Events

To evaluate the difference between the impact of scientific and nonscientific events on health beliefs, we conducted a Kruskal-Wallis test. The Kruskal-Wallis test was chosen since there was no reasonable assumed distribution for the influence

of the two types of events, and the two groups being compared had different sample sizes. We collected events associated with HCQ and CQ on the internet during the study period with no exclusion criteria. All the events were classified as scientific events if they were based on scientific evidence or endorsed by authorities, while all remaining events were treated as nonscientific events. To quantify the influences, for each event, we calculated the sum of the number of tweets that expressed perceived benefits or perceived barriers regarding HCQ and CQ on the day of the event and the day after.

Results

Data and Machine Learning Classifiers

The data set contained identifiers for 104,512,658 unique tweets, of which 92,687,660 were still available upon extraction. After applying the language filter, our final set for analysis consisted of 51,792,817 English tweets. The Cohen κ score for interrater reliability of data annotation was 0.94 for identifying whether a given tweet was HBM related and was around 0.9 for the annotation of all four individual HBM constructs (see Table 2). Random forest was found to be the best-performing classifier for the HBM-related classification and three of the HBM constructs (ie, perceived susceptibility, perceived benefits, and perceived barriers), while the passive-aggressive classifier was found to be the most suitable choice for classifying whether a tweet indicated perceived severity. The AUROC curves for HBM-related and the two disease-related constructs were all above 0.9, while AUROC curves for the two treatment-related constructs were around 0.86 (see Table 2).

After classification, 5,585,780 tweets were HBM related, among which 3,058,121 (54.75%) tweets expressed perceived susceptibility of COVID-19, 2,239,038 (40.08%) tweets expressed perceived severity of COVID-19, 211,374 (0.04%) tweets expressed perceived benefits of HCQ or CQ, and 190,839 (0.03%) tweets expressed perceived barriers toward HCQ or CQ. To further ensure the validity of the classification, we performed additional spot checks on the final results; examples can be found in Multimedia Appendix 1.

Table 2. Performance of machine learning classifiers.

Construct	Cohen κ[a]	Classifier[b]	AUROC[c] curve	Accuracy	Precision[d]	Recall[e]	F1 score[f]
Perceived susceptibility	0.92	Random forest	0.97	0.94	0.92	0.85	0.88
Perceived severity	0.88	Passive-aggressive	0.92	0.90	0.88	0.77	0.81
Perceived benefits	0.92	Random forest	0.87	0.79	0.78	0.78	0.78
Perceived barriers	0.92	Random forest	0.86	0.77	0.77	0.77	0.77
HBM[g] related	0.94	Random forest	0.90	0.84	0.84	0.84	0.84

[a]The Cohen κ coefficient for interrater reliability of annotation.

[b]The machine learning classifier selected by the best performance.

[c]AUROC: area under the receiver operating characteristic.

[d]Macro-averaged precision.

[e]Macro-averaged recall.

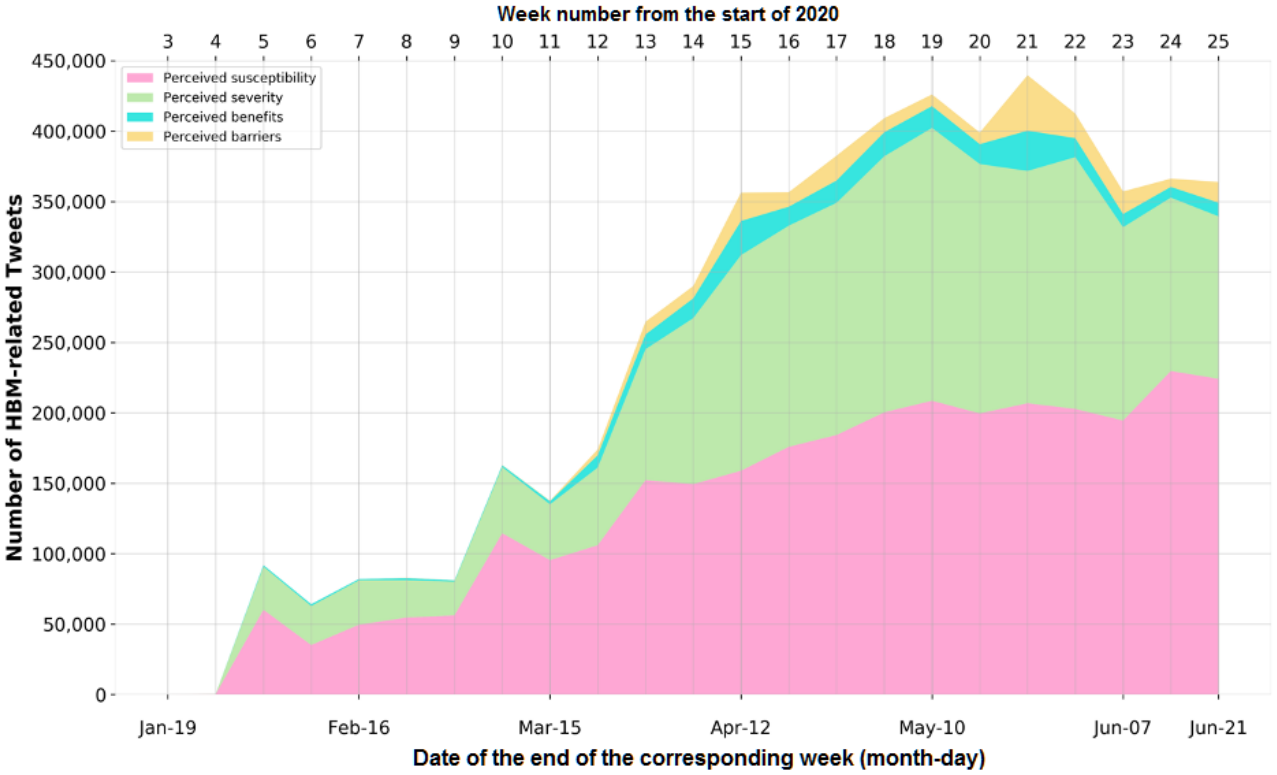
[f]Macro-averaged F1 score.

[g]HBM: health belief model.

The Overall Trend of Health Beliefs in Tweets

The visualization of the overall trend of health beliefs is shown in Figure 3, with the number of tweets that fell into each core construct of the HBM. Each construct was displayed in a different color chronologically, starting from the third week of 2020, and stacked together to show the total number of HBM-related tweets. A dramatic increase can be observed from January to June, which indicates an increasing number of discussions regarding personal health beliefs. The R_0 was 7.62 for the users who tweeted about health beliefs in our data.

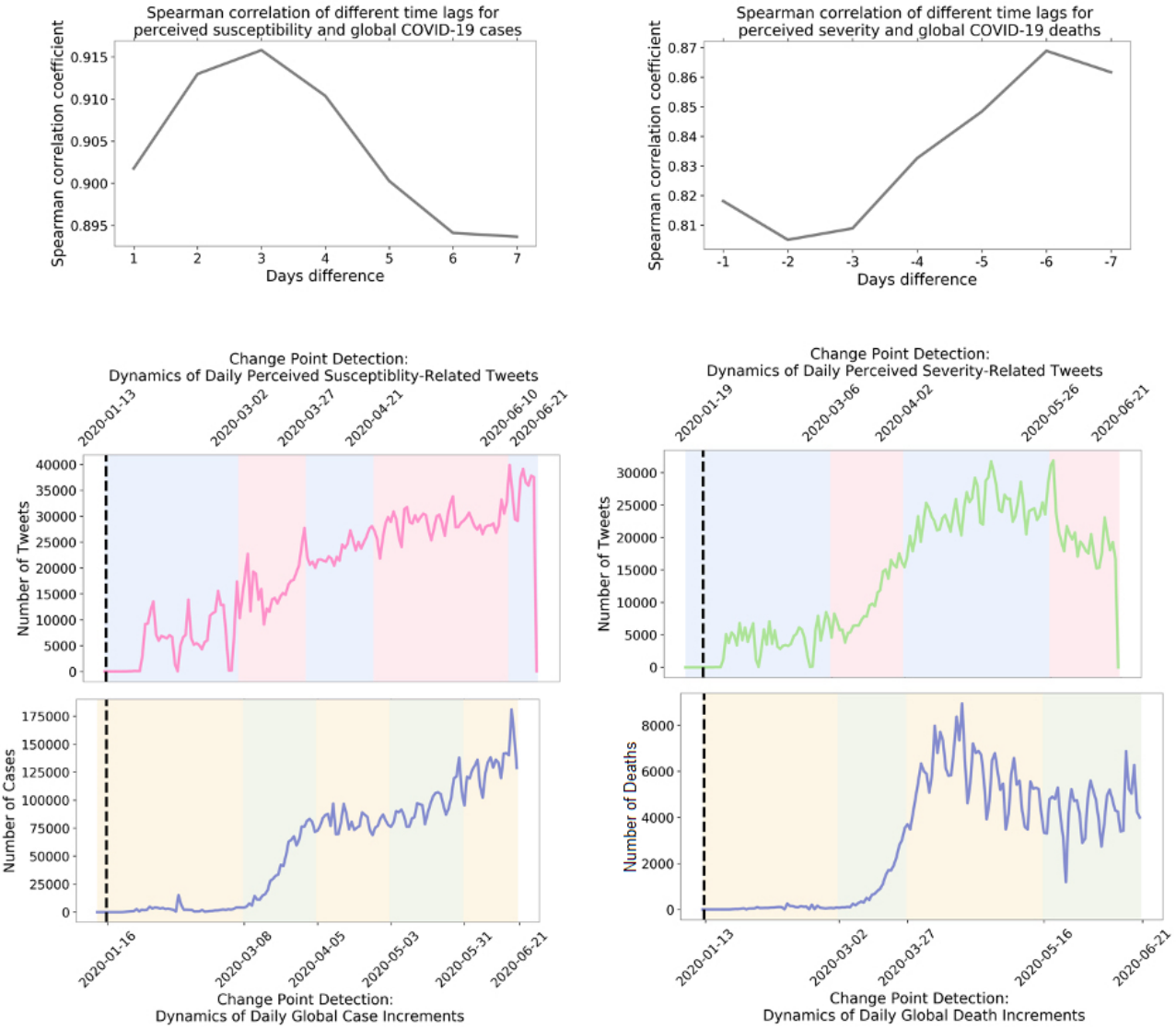
Figure 3. Stacked area chart for the four core constructs of the health belief model (HBM) from January 19 to June 21, 2020.



The Trend of Health Belief Toward the Disease

The left panel of Figure 4 displays the strongest correlation (ie, 0.92) between perceived susceptibility–related tweets and the global case increment when imposing a 3-day lag (ie, moving the trend of COVID-19 cases 3 days forward so that the number of perceived susceptibility tweets on January 13, 2020, will be aligned with the number of COVID-19 cases on January 16, 2020). The patterns detected by the change point analysis depicted by color in Figure 4 also show similarities within the pair. For the second pair (ie, perceived severity and COVID-19 death trend in the right-hand panel of Figure 4), the strongest correlation was found at –6 days (ie, 0.87), which indicates that changes in the perceived severity were lagging the actual death dynamics by 6 days (ie, the strongest correlation was found when moving the death trend 6 days backward). The change point analysis unraveled similar patterns between the trends of perceived severity and COVID-19 deaths.

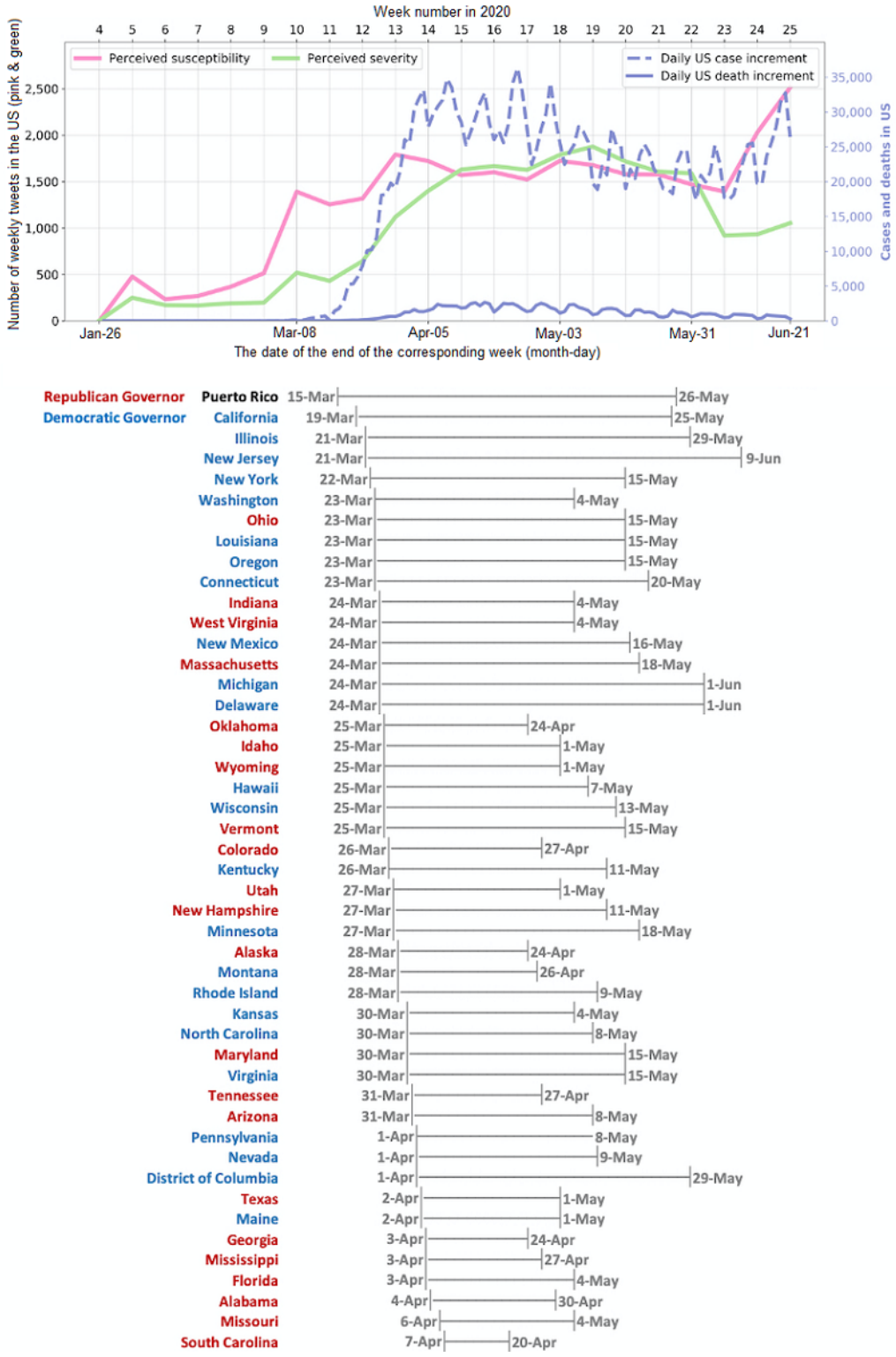
Figure 4. Correlation between perceived susceptibility–related tweets and COVID-19 case dynamics. Each pair of the lower four graphs are staggered according to the time differences that achieve the highest correlation in the top graphs: 3 days and –6 days, respectively. The pink and blue as well as the yellow and green shades depict the change points detected from the change point analysis.



The Effect of Interventions

We visualized the trends of perceived susceptibility and perceived severity along with the daily case and death dynamics in the United States in Figure 5. Meanwhile, lockdown information for each state is also listed by the timeline. Decisions made by Republican or Democratic governors are colored red and blue, respectively. The Republican states South Dakota, North Dakota, Iowa, Nebraska, and Arkansas did not announce official lockdowns and are not included in this figure. The official documents of lockdown and reopen decisions for each state are listed in Multimedia Appendix 1, Table S6.

Figure 5. Dynamics of health beliefs related to COVID-19 and the trend of case and death fluctuation in the United States with lockdown status. The lower half of the figure shows the official lockdown circumstances in each US state by each governor. The lines corresponding to each state represent the start and end date of the official lockdowns. The Republican states South Dakota, North Dakota, Iowa, Nebraska, and Arkansas did not announce official lockdowns and are not shown in this figure. The full reference for each state can be found in Multimedia Appendix 1, Table S6.



News in the Top Topics

Top 10 topics, which where all changed to lowercase, according to the tf-idf scores are shown in Figure 6, where a darker shade of the cell indicates a higher tf-idf score. The featured phrases that were closely related to the news during the corresponding time periods are highlighted in purple.

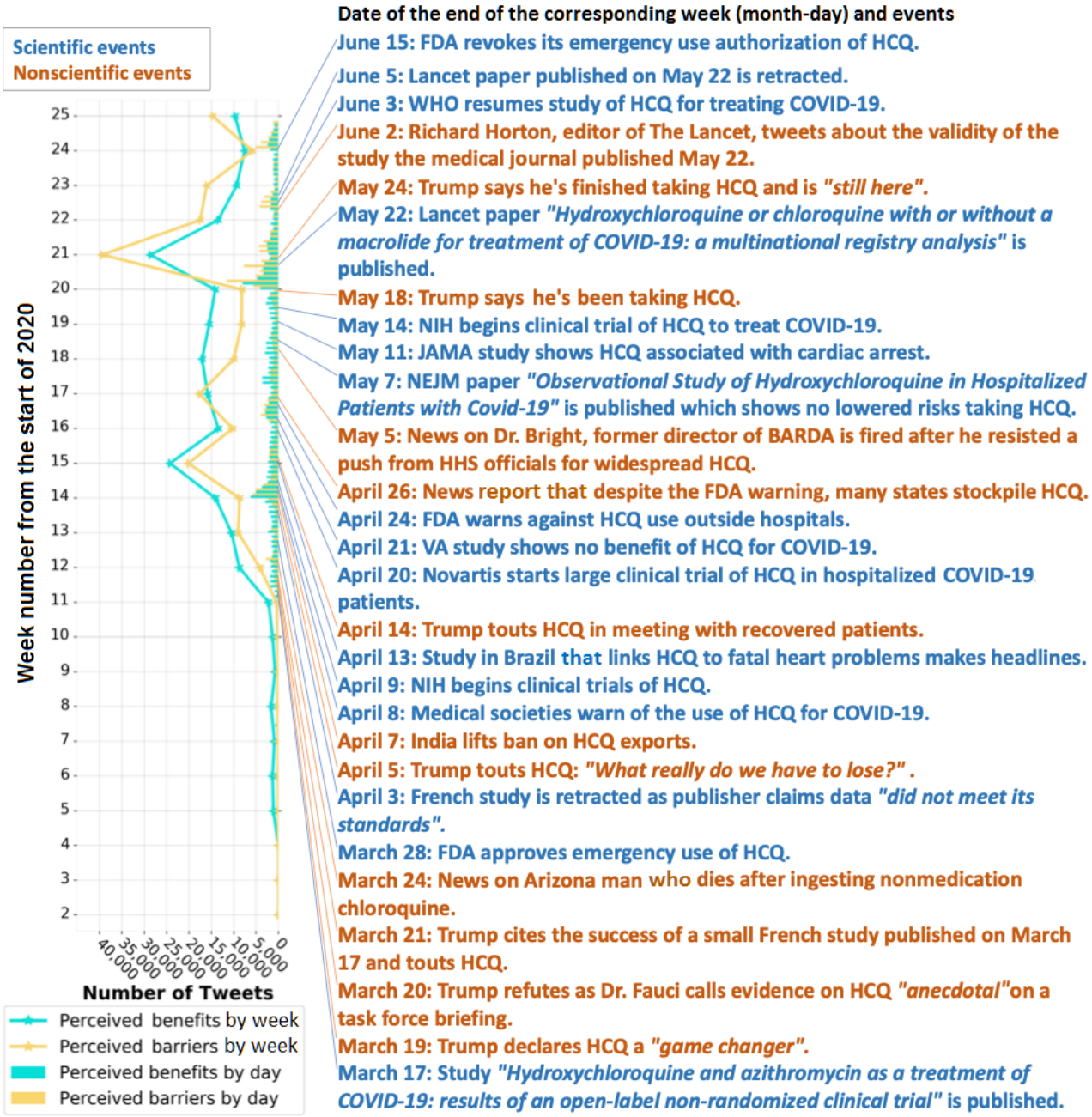
Figure 6. Top 10 topics of each week. Top phrases for each week are organized horizontally in each row. The blue shade in each cell indicates the term frequency–inverse document frequency (tf-idf) score of the phrase; the higher the tf-idf score, the darker the shade. The phrases that were likely associated with news are highlighted in purple.

Week	Data interval	1	2	3	4	5	6	7	8	9	10
2	Jan-06~Jan-12	advisory chinese	experts search	low experts	via tldr	viral nucleic	difficult confirm	search answers	notify health	weeks returning	outbreak ask
3	Jan-13~Jan-19	closed known	market infection	turn lethal	lethal likely	spread seafood	coronavirus rpts	ncov japan	report laboratory	major transport	dies novel
4	Jan-20~Jan-26	spreads multiple	france confirms	confirmed chicago	texas student	publishes early	warns grave	alarming consequences	consequenc es	points alarming	exposed lancet
5	Jan-27~Feb-02	outside china	philippines	spreads multiple	germany	confirmed uk	spreads regions	russia	outbreak spreads	rises outbreak	jumps outbreak
6	Feb-03~Feb-09	cruise	ship	cruise ship	li	wenliang	li wenliang	jumps outbreak	chinese rage	whistleblo wer	death coronavirus
7	Feb-10~Feb-16	ship	cruise	cruise ship	covid	reports new	spike	trump	seemed leveling	outside china	princess
8	Feb-17~Feb-23	iran	italy	covid	ship	cruise	cruise ship	outside china	spike	reports new	diamond princess
9	Feb-24~Mar-01	italy	iran	covid	trump	outside china	many cases	china real	coronavirus many	countries affected	updates number
10	Mar-02~Mar-08	covid	italy	iran	trump	york	new york	covid cases	cases covid	cruise	declares
11	Mar-09~Mar-15	covid	italy	trump	covid cases	cases covid	iran	coronavirus covid	spain	spread covid	community
12	Mar-16~Mar-22	covid	italy	trump	covid cases	chloroquine	cases covid	covid patients	spread covid	coronavirus pandemic	old
13	Mar-23~Mar-29	covid	trump	covid cases	italy	covid patients	chloroquine	cases covid	york	new york	hydroxychl oroquine
14	Mar-30~Apr-05	covid	trump	covid cases	covid patients	hydroxychlo roquine	italy	york	new york	spain	cases covid
15	Apr-06~Apr-12	covid	trump	hydroxychl oroquine	covid cases	covid patients	covid deaths	york	new york	italy	hours
16	Apr-13~Apr-19	covid	trump	covid cases	covid patients	covid deaths	hydroxychlo roquine	cases covid	homes	york	new york
17	Apr-20~Apr-26	covid	trump	covid cases	covid patients	hydroxychlo roquine	covid deaths	cases covid	york	new york	nursing
18	Apr-27~May-03	covid	covid cases	trump	covid patients	covid deaths	remdesivir	cases covid	homes	nursing	new deaths
19	May-04~May-10	covid	trump	covid cases	covid patients	covid deaths	highest	nursing	homes	cases covid	new deaths
20	May-11~May-17	covid	covid cases	trump	covid patients	nursing	covid deaths	homes	nursing homes	hydroxychlo roquine	cases covid
21	May-18~May-24	covid	trump	hydroxychlo roquine	covid cases	covid patients	nursing	homes	covid deaths	nursing homes	cases covid
22	May-25~May-31	covid	covid cases	trump	covid patients	covid deaths	hydroxychlo roquine	nursing	homes	nursing homes	highest
23	Jun-01~Jun-07	covid	covid cases	covid patients	trump	hydroxychlo roquine	covid deaths	new covid	george floyd	hours	spike
24	Jun-08~Jun-14	covid	covid cases	trump	covid patients	spike	active cases	highest	new covid	cases covid	covid deaths
25	Jun-15~Jun-21	covid	covid cases	trump	dexametha sone	florida	covid patients	spike	record	new covid	cases covid

The Influence of Scientific and Nonscientific Events

The list of events that we collected is shown on the right-hand side of Figure 7, while the trends of perceived benefits and barriers are shown on the left-hand side. We observed that both scientific and nonscientific events were associated with fluctuations in health beliefs. The scales of the fluctuations observed varied over time. There were more nonscientific events around the two most massive spikes, but scientific events were majorly distributed along the timeline where many gentle fluctuations could be found. The Kruskal-Wallis test showed no significant difference between the influence of scientific and nonscientific events for both perceived benefits and barriers (H=0.078, P=.78; and H=0.002, P=.92, respectively). Full references for each event can be found in Multimedia Appendix 1, Table S7.

Figure 7. Dynamics of health beliefs related to hydroxychloroquine (HCQ) and chloroquine (CQ) with correlated scientific and nonscientific events. Scientific events are those that have concrete scientific evidence or are endorsed by authorities, while nonscientific events account for the rest. The full reference for each event can be found in Multimedia Appendix 1, Table S7. BARDA: Biomedical Advanced Research and Development Authority; FDA: US Food and Drug Administration; HHS: United States Department of Health and Human Services; JAMA: The Journal of the American Medical Association; NEJM: The New England Journal of Medicine; NIH: National Institutes of Health; VA: United States Department of Veterans Affairs; WHO: World Health Organization.



Discussion

Principal Findings

Through the utilization of natural language processing (NLP) and machine learning, we employed the HBM to identify tweets associated with health beliefs. Through further evaluation of HBM-related tweets, our findings demonstrated that trends in health beliefs were correlated with dynamics in positive case and mortality rates. Additionally, we observed a decline in perceived disease susceptibility during government-issued lockdowns, while perceived severity appeared unaltered. Lastly, our study identified top news events, scientific and nonscientific, that may play a role in altering health beliefs. These findings lay the groundwork to better understand how the general public's COVID-19-related health beliefs are influenced by case and mortality rates, government policies, current news, and significant events. Through careful study of these observations, we may better implement management strategies to combat the pandemic and the infodemic.

In commonly used models for infectious diseases, infection is considered to be spreading in a population when R_0 is greater than 1, and the epidemic is harder to control with a larger value of R_0. Therefore, given the R_0 of 7.62, it is reasonable to conclude that an infodemic is ongoing in our study population.

It is interesting that health beliefs involving perceived susceptibility increased in advance of the actual evolvement of the pandemic. Because we observed the basic reproduction number R_0 of 7.62, suggestive of an infodemic, these findings may suggest that the volume of information regarding COVID-19 affects Twitter users' perspectives regarding the risk of infection. In the early stages of the pandemic, before mortalities were observed, it is possible that less severity was assumed. Over time, perceived severity may have increased as the number of deaths cumulated. Strong correlations between perceived susceptibility and perceived severity regarding the case and death dynamics may suggest that the ongoing situation of the pandemic is a significant impact factor affecting health beliefs.

From the line chart in Figure 5, we observed a dramatic increase in daily cases between week 11 and week 14. There was also an upward trend in perceived susceptibility starting from week 11, which began decreasing by week 13. This phenomenon is interesting when we take the lockdown situation into consideration, as starting from week 13 was when most of the states were under the government-issued lockdown. Thus, official interventions were observed to potentially mitigate the general public's perceived susceptibility of COVID-19. Meanwhile, we saw the growth of the number of confirmed cases slowing down during this same period. Interestingly, we failed to see a decline in perceived severity even when almost all the states were under quarantine. Previously, we showed that perceived severity was found to most strongly correlate with mortality; thus, it is reasonable that lockdown policies did not ease such health concerns, perhaps owing to the fact that while lockdowns slow down the spread of infection, they do not offer complete protection, especially in the absence of viable medications or treatment strategies.

As shown in Figure 6, for the first 3 to 4 weeks, topics predominately covered confirmed cases worldwide when the global pandemic was not yet affirmed by the authorities. In the following weeks, the terms *cruise ship* [29] and *li wenliang* [30] came into the spotlight. In early February, a large and notable cluster of COVID-19 cases occurred on the Diamond Princess cruise ship. Dr Wenliang Li, the Chinese doctor who tried to raise the alarm about a possible outbreak of a disease that resembled SARS-CoV-2 in Wuhan, China, died of the infection on February 7 , 2020. During weeks 8 and 9, when the COVID-19 outbreak heightened in Italy [31] and Iran [32], topics related to these countries began trending. On March 18, 2020 (ie, week 12), then President Trump announced that he was taking HCQ as prophylaxis for COVID-19 [33] and triggered massive discussions. In fact, starting at week 13, discussions involving HCQ and CQ began to dominate. Lastly, we were initially surprised to observe other topics like *george floyd* [34] in the health belief–related tweets. However, this topic is related to many events where people gathered that happened while many states were still under lockdown, possibly provoking health concerns. Through this analysis, we suspect that the news from all sources may penetrate into discussions regarding health beliefs and, thus, may influence health beliefs. Therefore, the news that we consume every day may inadvertently be a substantial factor that affects our health beliefs, which may contribute to and even exacerbate the infodemic on social media.

We observed that speeches by politicians could have dramatic impacts on the health beliefs of the general public who read the news. However, politicians' speeches do not necessarily recapitulate scientific facts or evidence and could sometimes be misleading [35]. Thus, we expect to rely more on scientific sources, such as publications with scientific evidence or announcements made by health authorities, for more accurate and reliable information regarding the pandemic. However, it is uncertain whether scientific events or nonscientific events have a more profound influence on altering the health beliefs of the general public.

The results from the Kruskal-Wallis test imply that scientific events and nonscientific events did not significantly differ from one another in regard to their effect on health beliefs within the given period (ie, January 6 to June 21, 2020). We found it surprising that scientific events did not appear to be significantly associated with altering the health beliefs toward potential treatments in our data set. This might be due to the public's distrust in science arising from the many uncertainties involving the pandemic or the instances of being delivered conflicting information, such as "Don't wear masks" to "Wear masks all the time." To better cope with COVID-19 circumstances, everyone in society, online and offline, should be aware of the overabundance of information and its potential impact on health beliefs. Thus, it is essential to be prudent to screen the authenticity of each piece of information.

Limitations and Future Work

We have identified some limitations in this study. The tweets analyzed in this study covered the English language, and there might be divergences across different languages that were not addressed. Although English tweets constitute the largest proportion among all the tweets, the number of tweets in other languages or undefined languages are still considerable [36]. We hope to expand the analysis to a multilingual setting in future work. Additionally, although we did not cover every potential treatment at this stage, our framework is extensible to assess the influence on health beliefs of additional treatments or interventions, such as vaccines. In fact, we plan to apply similar approaches to investigate health beliefs in COVID-19 vaccines once they are available to the general public. Furthermore, we likely have not considered other factors that may contribute to alterations in health beliefs.

This analysis used data extracted from one social media platform, Twitter, which may also introduce bias. Users' health beliefs may not represent those of the entire population, since not everyone uses Twitter. More social media platforms will be incorporated in future work, such as Facebook, Instagram, and Reddit. Additionally, it would be interesting to compare our crowdsourcing results with health beliefs obtained through hospital-administrated surveys from patients with COVID-19 and their caretakers.

Technically, this study employed the very classic text classification methods, which used a combination of the bag-of-words model and machine learning classifiers. Yet, the

experiments showed that they worked well (ie, AUROC curve over 0.9) on the given data. Deep learning architectures were not discussed in this study, mainly because there is no guarantee that deep learning models always work better than simple machine learning classifiers. Meanwhile, deep learning models bear higher technical barriers, which compromise the accessibility for people from other domains. Deep learning models are also known to demand considerable energy [37], so we were also trying to trade off the energy-performance balance. However, it is definitely worth a whole other study to discuss various NLP techniques for the classification task. For future studies, we are also interested in investigating the performance of various NLP techniques on the current text classification tasks.

Conclusions

Our data suggest that we are not only fighting a pandemic but also an infodemic. The excessive information disseminated on social media platforms and other sources is closely related to the dynamics of the general public's health beliefs. The dynamics of the pandemic, news, scientific and nonscientific events, and even the related tweets already published on social media platforms may influence the health beliefs of the general public on social media to some extent. Our findings provide clues and evidence for more effective management of the infodemic associated with the COVID-19 pandemic.

Acknowledgments

HW, YLi, and YLuo conceived of the presented idea. HW, YLi, and MH contributed to the data annotation and validation. HW carried out all the experiments and conducted all the data analyses. HW wrote the manuscript and designed the tables and figures with support from MH and YLi. AN provided clinical insights and contributed to the results interpretation. YLuo advised and supervised the entire project. We thank Twitter for providing the API for developers to retrieve tweets for this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information.
[DOCX File , 70 KB - jmir_v23i2e26302_app1.docx]

References

1. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
3. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733 [FREE Full text] [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
4. Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ [accessed 2020-10-17]
5. Zarocostas J. How to fight an infodemic. Lancet 2020 Feb 29;395(10225):676 [FREE Full text] [doi: 10.1016/S0140-6736(20)30461-X] [Medline: 32113495]
6. Eysenbach G. Infodemiology: The epidemiology of (mis)information. Am J Med 2002 Dec 15;113(9):763-765. [doi: 10.1016/s0002-9343(02)01473-0] [Medline: 12517369]
7. 1st WHO Infodemiology Conference. World Health Organization. 2020. URL: https://www.who.int/news-room/events/detail/2020/06/30/default-calendar/1st-who-infodemiology-conference [accessed 2020-10-17]
8. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
9. Eysenbach G. How to fight an infodemic: The four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
10. Hua J, Shaw R. Corona virus (COVID-19) "infodemic" and emerging issues through a data lens: The case of China. Int J Environ Res Public Health 2020 Mar 30;17(7):2309 [FREE Full text] [doi: 10.3390/ijerph17072309] [Medline: 32235433]
11. Janz NK, Becker MH. The health belief model: A decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
12. Rosenstock IM. Historical origins of the health belief model. Health Educ Monogr 1974 Dec 01;2(4):328-335. [doi: 10.1177/109019817400200403]

13. Strecher V, Rosenstock I. The health belief model. In: Baum A, Newman S, Weinman J, West R, McManus C, editors. Cambridge Handbook of Psychology, Health and Medicine. Cambridge, UK: Cambridge University Press; 1997:113-116.
14. Rosenstock IM, Strecher VJ, Becker MH. The health belief model and HIV risk behavior change. In: DiClemente RJ, Peterson JL, editors. Preventing AIDS: Theories and Methods of Behavioral Interventions. Boston, MA: Springer; 1994:5-24.
15. Du J, Cunningham RM, Xiang Y, Li F, Jia Y, Boom JA, et al. Leveraging deep learning to understand health beliefs about the human papillomavirus vaccine from social media. NPJ Digit Med 2019;2:27 [FREE Full text] [doi: 10.1038/s41746-019-0102-4] [Medline: 31304374]
16. Wardle J, Haase AM, Steptoe A, Nillapun M, Jonwutiwes K, Bellisle F. Gender differences in food choice: The contribution of health beliefs and dieting. Ann Behav Med 2004 Apr;27(2):107-116. [doi: 10.1207/s15324796abm2702_5] [Medline: 15053018]
17. Czeisler M, Tynan MA, Howard ME, Honeycutt S, Fulmer EB, Kidder DP, et al. Public attitudes, behaviors, and beliefs related to COVID-19, stay-at-home orders, nonessential business closures, and public health guidance - United States, New York City, and Los Angeles, May 5-12, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 19;69(24):751-758 [FREE Full text] [doi: 10.15585/mmwr.mm6924e1] [Medline: 32555138]
18. Guidry JP, Carlyle KE, LaRose JG, Perrin P, Messner M, Ryan M. Using the health belief model to analyze Instagram posts about Zika for public health communications. Emerg Infect Dis 2019 Jan;25(1):179-180 [FREE Full text] [doi: 10.3201/eid2501.180824] [Medline: 30561302]
19. Sesagiri Raamkumar A, Tan SG, Wee HL. Use of health belief model-based deep learning classifiers for COVID-19 social media content to examine public perceptions of physical distancing: Model development and case study. JMIR Public Health Surveill 2020 Jul 14;6(3):e20493 [FREE Full text] [doi: 10.2196/20493] [Medline: 32540840]
20. Carico RRJ, Sheppard J, Thomas CB. Community pharmacists and communication in the time of COVID-19: Applying the health belief model. Res Social Adm Pharm 2021 Jan;17(1):1984-1987 [FREE Full text] [doi: 10.1016/j.sapharm.2020.03.017] [Medline: 32247680]
21. Banda J, Tekumalla R, Wang G, Yu J, Liu T, Ding Y, et al. A large-scale COVID-19 Twitter chatter dataset for open scientific research--An international collaboration. arXiv. Preprint posted online on April 7, 2020. [FREE Full text]
22. Tekumalla R, Banda JM. Social Media Mining Toolkit (SMMT). Genomics Inform 2020 Jun;18(2):e16. [doi: 10.5808/gi.2020.18.2.e16]
23. McHugh ML. Interrater reliability: The kappa statistic. Biochem Med 2012;22(3):276-282. [doi: 10.11613/bm.2012.031]
24. Wang H. Understanding COVID-19 health beliefs using tweets. GitHub. 2020. URL: https://github.com/HanyinWang/CovidHealthBeliefTweets [accessed 2021-02-09]
25. Bird S, Klein E, Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. Sebastopol, CA: O'Reilly Media; 2009.
26. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
27. Bailey NT. The Mathematical Theory of Infectious Diseases and its Applications. London, UK: Charles Griffin & Company; 1975.
28. Truong C, Oudre L, Vayatis N. Selective review of offline change point detection methods. Signal Processing 2020 Feb;167:107299. [doi: 10.1016/j.sigpro.2019.107299]
29. Moriarty L, Plucinski MM, Marston BJ, Kurbatova EV, Knust B, Murray EL, CDC Cruise Ship Response Team, California Department of Public Health COVID-19 Team, Solano County COVID-19 Team. Public health responses to COVID-19 outbreaks on cruise ships - Worldwide, February-March 2020. MMWR Morb Mortal Wkly Rep 2020 Mar 27;69(12):347-352 [FREE Full text] [doi: 10.15585/mmwr.mm6912e3] [Medline: 32214086]
30. Green A. Li Wenliang. Lancet 2020 Feb 18;395(10225):682 [FREE Full text] [doi: 10.1016/s0140-6736(20)30382-2]
31. Remuzzi A, Remuzzi G. COVID-19 and Italy: What next? Lancet 2020 Apr 11;395(10231):1225-1228 [FREE Full text] [doi: 10.1016/S0140-6736(20)30627-9] [Medline: 32178769]
32. Abdi M. Coronavirus disease 2019 (COVID-19) outbreak in Iran: Actions and problems. Infect Control Hosp Epidemiol 2020 Jun;41(6):754-755 [FREE Full text] [doi: 10.1017/ice.2020.86] [Medline: 32192541]
33. Lovelace Jr B, Breuninger K. Trump says he takes hydroxychloroquine to prevent coronavirus infection even though it's an unproven treatment. CNBC. 2020 May 18. URL: https://www.cnbc.com/2020/05/18/trump-says-he-takes-hydroxychloroquine-to-prevent-coronavirus-infection.html [accessed 2020-08-27]
34. George Floyd protests. Wikipedia. 2020. URL: https://en.wikipedia.org/wiki/George_Floyd_protests [accessed 2020-08-27]
35. Lewandowsky S, Jetter M, Ecker U. Using the president's tweets to understand political diversion in the age of social media. Nat Commun 2020 Nov 10;11(1):5764 [FREE Full text] [doi: 10.1038/s41467-020-19644-6] [Medline: 33173060]
36. Pobiruchin M, Zowalla R, Wiesner M. Temporal and location variations, and link categories for the dissemination of COVID-19-related information on Twitter during the SARS-CoV-2 outbreak in Europe: Infoveillance study. J Med Internet Res 2020 Aug 28;22(8):e19629 [FREE Full text] [doi: 10.2196/19629] [Medline: 32790641]
37. Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019 Presented at: 57th Annual Meeting of the

Association for Computational Linguistics; July 28-August 2, 2019; Florence, Italy p. 3645-3650 URL: https://www.aclweb.org/anthology/P19-1355.pdf [doi: 10.18653/v1/p19-1355]

Abbreviations

API: application programming interface
AUROC: area under the receiver operating characteristic
CQ: chloroquine
HBM: health belief model
HCQ: hydroxychloroquine
HPV: human papillomavirus
NLP: natural language processing
SIR: susceptible-infectious-recovered
tf-idf: term frequency–inverse document frequency
WHO: World Health Organization

Edited by R Kukafka, C Basch; submitted 06.12.20; peer-reviewed by J Banda Orozco, R Zowalla; comments to author 22.12.20; revised version received 08.01.21; accepted 31.01.21; published 22.02.21.

Please cite as:
Wang H, Li Y, Hutch M, Naidech A, Luo Y
Using Tweets to Understand How COVID-19–Related Health Beliefs Are Affected in the Age of Social Media: Twitter Data Analysis Study
J Med Internet Res 2021;23(2):e26302
URL: https://www.jmir.org/2021/2/e26302
doi:10.2196/26302
PMID:33529155

Original Paper

Establishing Classifiers With Clinical Laboratory Indicators to Distinguish COVID-19 From Community-Acquired Pneumonia: Retrospective Cohort Study

Wanfa Dai[1*], MD; Pei-Feng Ke[2,3*], MPH; Zhen-Zhen Li[4], BSc; Qi-Zhen Zhuang[4], BSc; Wei Huang[1], MD; Yi Wang[2,4], MPhil; Yujuan Xiong[2,3*], PhD; Xian-Zhang Huang[2,3*], PhD

[1]Department of Respiration, Gong An County People's Hospital, Jingzhou, China

[2]Department of Laboratory Medicine, The Second Affiliated Hospital, Guangzhou University of Chinese Medicine, Guangzhou, China

[3]Guangdong Provincial Key Laboratory of Research on Emergency in Traditional Chinese Medicine, Guangzhou, China

[4]Second Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Xian-Zhang Huang, PhD
Department of Laboratory Medicine
The Second Affiliated Hospital
Guangzhou University of Chinese Medicine
111 Dade Rd
Guangzhou, 510210
China
Phone: 86 020 81887233 ext 35362
Email: huangxz020@gzucm.edu.cn

Abstract

Background: The initial symptoms of patients with COVID-19 are very much like those of patients with community-acquired pneumonia (CAP); it is difficult to distinguish COVID-19 from CAP with clinical symptoms and imaging examination.

Objective: The objective of our study was to construct an effective model for the early identification of COVID-19 that would also distinguish it from CAP.

Methods: The clinical laboratory indicators (CLIs) of 61 COVID-19 patients and 60 CAP patients were analyzed retrospectively. Random combinations of various CLIs (ie, CLI combinations) were utilized to establish COVID-19 versus CAP classifiers with machine learning algorithms, including random forest classifier (RFC), logistic regression classifier, and gradient boosting classifier (GBC). The performance of the classifiers was assessed by calculating the area under the receiver operating characteristic curve (AUROC) and recall rate in COVID-19 prediction using the test data set.

Results: The classifiers that were constructed with three algorithms from 43 CLI combinations showed high performance (recall rate >0.9 and AUROC >0.85) in COVID-19 prediction for the test data set. Among the high-performance classifiers, several CLIs showed a high usage rate; these included procalcitonin (PCT), mean corpuscular hemoglobin concentration (MCHC), uric acid, albumin, albumin to globulin ratio (AGR), neutrophil count, red blood cell (RBC) count, monocyte count, basophil count, and white blood cell (WBC) count. They also had high feature importance except for basophil count. The feature combination (FC) of PCT, AGR, uric acid, WBC count, neutrophil count, basophil count, RBC count, and MCHC was the representative one among the nine FCs used to construct the classifiers with an AUROC equal to 1.0 when using the RFC or GBC algorithms. Replacing any CLI in these FCs would lead to a significant reduction in the performance of the classifiers that were built with them.

Conclusions: The classifiers constructed with only a few specific CLIs could efficiently distinguish COVID-19 from CAP, which could help clinicians perform early isolation and centralized management of COVID-19 patients.

(J Med Internet Res 2021;23(2):e23390) doi:10.2196/23390

KEYWORDS

COVID-19; clinical laboratory indicators; community-acquired pneumonia; classifier; classification algorithm

Introduction

COVID-19 caused by SARS-CoV-2 infection, which was discovered in early December 2019, has become a global pandemic. As of August 3, 2020, COVID-19 has become widespread in 215 countries, areas, or territories worldwide; it has caused infection in more than 17.9 million people and has resulted in the deaths of more than 686,000 people [1]. The World Health Organization has stated that the spread of COVID-19 may be impeded by early detection, isolation, and the implementation of a robust health care system [2,3]. Nevertheless, the published data indicate that the initial symptoms of COVID-19 in patients are very similar to those in patients with the common cold or influenza. COVID-19 patients exhibit different clinical symptoms, and some of them do not have any symptoms [4-7]. sars-cov-2 infection has a long incubation period, with a median incubation period of 5 to 7 days, which is the chief risk factor for community infection [6,8]. Community-acquired pneumonia (CAP) and COVID-19 have similar clinical and imaging features, but their treatment and infectivity are very dissimilar. Distinguishing COVID-19 from CAP is very important to prevent the spread of COVID-19 and to provide specific treatment.

Some characteristic spectra demonstrated by clinical laboratory indicators (CLIs) of COVID-19 patients have been utilized as auxiliary clues for diagnosis [9]. Previous studies have demonstrated that increased procalcitonin (PCT), lymphocytopenia, and thrombin activation can all be utilized as auxiliary diagnostic indicators of COVID-19 and poor prognostic factors [9-11]. However, they are also correlated with CAP [12-15]. Thus, in accordance with the changes in these indicators, it is impossible to differentiate COVID-19 from CAP. The changes in the neutrophil to lymphocyte ratio, the peak platelet to lymphocyte ratio, lactate dehydrogenase (LDH), C-reactive protein (CRP), and interleukin-6 (IL-6) are considered to be associated with the progression and prognosis of COVID-19 [9], but using the information from the CLIs to give clinicians correct guidance is still a great challenge.

Classifiers established by machine learning (ML) algorithms based on various clinical features, biomarkers, and CLIs are increasingly widely utilized in disease diagnosis and risk prediction [16]. During the COVID-19 pandemic, ML was also widely used to predict, classify, assess, track, and control the spread of SARS-CoV-2 [17,18]. ML can improve diagnostic performance compared with hand-selected biomarkers by selecting relevant biomarkers and more consistently capturing both their relative importance to prediction and their interactions among one another [19]. In this study, we used CLIs to build classifiers with different ML algorithms to distinguish COVID-19 patients from CAP patients; we found that only the feature combinations (FCs) with many specific CLIs rather than the FCs with the most significantly differential CLIs between the two groups could build high-performance classifiers (HPCs).

Methods

Collection of Patients' Electronic Medical Record Data

The electronic medical records of patients who were admitted to Gong An County People's Hospital, China, and diagnosed with COVID-19 or CAP from December 2019 to March 2020 were retrieved. The information regarding each patient's age, sex, clinical symptoms upon admission, medical history, epidemiological history, computed tomography (CT) imaging features, and CLIs were sorted out for retrospective analysis. Only the laboratory test results during admission were included. It was specified that all patients' data were to be kept confidential, and this data were only to be utilized for comprehensive analysis. No personal information about any patient was mentioned in the paper. This study was approved by the ethics committees from the Guangdong Provincial Hospital of Chinese Medicine (approval No. ZE2020-049-01) with a waiver of informed consent due to the retrospective nature of the study.

Data Description

Diagnosis and clinical classification of COVID-19 were performed according to the *Chinese Clinical Guidance for COVID-19 Pneumonia Diagnosis and Treatment (7th edition)* [20]. A total of 61 patients with COVID-19 and 60 patients with CAP were enrolled according to the discharge diagnosis on their electronic medical records. There were 3 mild, 47 common, 6 severe, and 5 critical types, which were categorized into two groups for further analysis as follows: COVID19-COM (3 mild and 47 common types) and COVID19-SV (6 severe and 5 critical types). They were matched by age and sex and did not significantly differ in terms of medical history. The main clinical symptoms between CAP and COVID-19 groups were not significantly different.

Primary Analysis

The descriptive analysis of all CLIs was performed between groups or subgroups. Between-group or between-subgroup differences were tested using the *statsmodels* module from Python (Python Software Foundation) [21]. The Student *t* test was performed when the distribution of the variables conformed to the normal distribution; otherwise, the Mann-Whitney *U* test was used. The chi-square test was used to detect differences in baseline data between two groups or subgroups. A value of $P<.05$ was considered to be significant.

Feature Selection and Data Preprocessing

The CLIs with a missing value ratio greater than 20% were excluded. Only the CLIs with a significant difference between the two groups were selected and used to generate 1,807,780 nonrepetitive random FCs, consisting of one to eight CLIs, by using the *combinations* iterator in the *itertools* module from Python [22]. Next, an FC was selected from the FC list one by one to form a new data sheet with the dependent variable (ie, disease type), and 1,807,780 new data sheets were eventually formed. For each new data sheet, the rows with missing values were removed. The remaining rows were then divided into *training_dataset* and *test_dataset* using scikit-learn, version 0.23.1 (*train_test_split* function with *test_size* = 0.25,

random_state = 0). The training data set was used to build the classifier, and the test data set was used to assess the performance. The feature values were standardized using the *StandardScaler* function in the scikit-learn module before constructing the logistic regression (LR) classifier.

Construction of Classifiers With ML Algorithms in the Scikit-Learn Module

Scikit-learn is a Python module integrating a wide range of state-of-the-art ML algorithms for medium-scale supervised and unsupervised problems [23]. The LR classifier, the random forest classifier (RFC), and the gradient boosting classifier (GBC) have been typically used to construct classifiers in prediction of disease risk, progression, prognosis, and so on [24]. The LR classifier in the *sklearn.linear_model* is also known as logit regression, maximum-entropy classification, or the log-linear classifier. In this model, the probabilities describing the possible outcomes of a single trial are modeled using a logistic function [24]. The RFC in the *sklearn.ensemble* module is one of the averaging algorithms in ensemble methods and is a perturb-and-combine technique specifically designed for trees. In the random forest algorithm, each tree in the ensemble is built from a sample drawn with replacement from the training data set. Furthermore, when splitting each node during the construction of a tree, the best split is found either from all input features or from a random subset of size setting with the parameter *max_features*. In practice, the variance reduction due to the introduction of randomness in the classifier construction is often significant, hence, yielding an overall better model [25,26]. The GBC algorithm, using the *sklearn.ensemble* function, is a boosting method, in which base estimators are built sequentially. To reduce the bias of the combined estimator, one has to combine several weak models to produce a powerful ensemble. The GBC algorithm builds an additive model in a forward stage-wise fashion, and it allows for the optimization of arbitrary differentiable loss functions [27,28].

In this study, the classifiers were respectively constructed using the LR classifier, RFC, and GBC in the scikit-learn module with the training data set. The model parameter settings were kept as default, except that *random_state* was modified to "0" for all models and *class_weight* was modified to "balanced" for the LR classifier and RFC models. The performance of the classifiers was evaluated with the test data set by calculating the recall rate (ie, sensitivity), specificity, accuracy, and area under the receiver operating characteristic curve (AUROC), using the *sklearn_metrics.recall_score*, *sklearn_metrics.precision_score*, *sklearn_metrics.accuracy_score*, and *sklearn_metrics.auc* functions, respectively. Gini importance was computed using the *feature_importance* function to measure the importance of each feature in the RFC and the GBC. The higher the Gini importance value, the more important the feature [29]. All the above analyses were performed in Python, version 3.7 (Python Software Foundation).

Results

Basic Characteristics of CAP Group and COVID-19 Group

No significant differences in age and sex were found between CAP and COVID-19 groups (see Table 1); however, the proportions of males in the CAP and COVID-19 groups were 55% (33/60) and 66% (40/61), respectively, and were higher than those of females in both groups. No significant difference in the medical history between the two groups (see Table 1) was observed. Also, no significant difference was found in the proportions of the main clinical symptoms between the two groups, such as fever, cough, fatigue, muscle soreness, and loss of appetite (see Table 1). The average hospitalization days for CAP patients were remarkably lower than those for COVID-19 patients (P<.001). In the CAP group, some patients with pulmonary CT also had imaging features that included patchy hyperdense shadow (11/60, 18%), ground-glass shadow (4/60, 7%), and fibrotic lesion (6/60, 10%). Nonetheless, the chief imaging features of pulmonary CT in the COVID-19 group were patchy hyperdense shadow (25/61, 41%) and ground-glass shadow (9/61, 15%), and many patients (7/61, 11%) had both patchy hyperdense shadow and ground-glass shadow (see Table 1). Among the 61 patients suffering from COVID-19, 3 (5%) had mild symptoms, 47 (77%) had common symptoms, 6 (10%) had severe symptoms, and 5 (8%) had critical symptoms. Fever and cough were the principal symptoms in the early stage of COVID-19, and these accounted for 70% (43/61) and 64% (39/61) of the cases, respectively (see Table 1). Among the CAP patients included in the analysis, no cases of death were found during hospitalization; however, 3 of the 5 (60%) severely ill patients in the COVID-19 group, who were aged 36, 49, and 74 years, died during hospitalization. The 36-year-old patient who died underwent interventricular septal repair in childhood.

Table 1. Comparison of baseline information between COVID-19 patients and community-acquired pneumonia (CAP) patients.

Baseline characteristic	CAP patients (n=60)	COVID-19 patients (n=61)	*P* value
Sex (male), n (%)	33 (55)	40 (66)	.27
Age (years), mean (SD)	55.72 (18.10)	50.23 (16.95)	.09
Hospitalization days, median (IQR)	9 (7-12)	21 (13-26)	<.001
Medical history, n (%)			
Hypertension	14 (23)	16 (26)	.83
Diabetes	2 (3)	6 (10)	.27
Liver disease	2 (3)	3 (5)	.99
Heart disease	3 (5)	5 (8)	.72
Exposure history	Unclear	54 (89)	N/A[a]
Familial aggregation infection[b]	Unclear	22 (36)	N/A
Initial symptoms, n (%)			
Fever	36 (60)	43 (70)	.26
Cough	44 (73)	39 (64)	.33
Myalgia	4 (7)	7 (11)	.53
Poor appetite	5 (8)	11 (18)	.18
Fatigue	33 (55)	24 (39)	.10
Days from onset of symptoms to admission, median (IQR)	Unrecorded	3 (1-7)	N/A
Imaging features, n (%)			
Patchy high-density opacity	11 (18)	25 (41)	.009
Ground-glass opacity	4 (7)	9 (15)	.24
Fibrotic lesion	6 (10)	3 (5)	.32
Patchy high-density opacity and ground-glass opacity	0 (0)	7 (11)	.01
Death cases, n (%)	0 (0)	3 (5)	N/A

[a]N/A: not applicable; groups could not be compared because there were no values for the CAP group.

[b]There were more than 2 cases of infection after aggregation with family members or relatives.

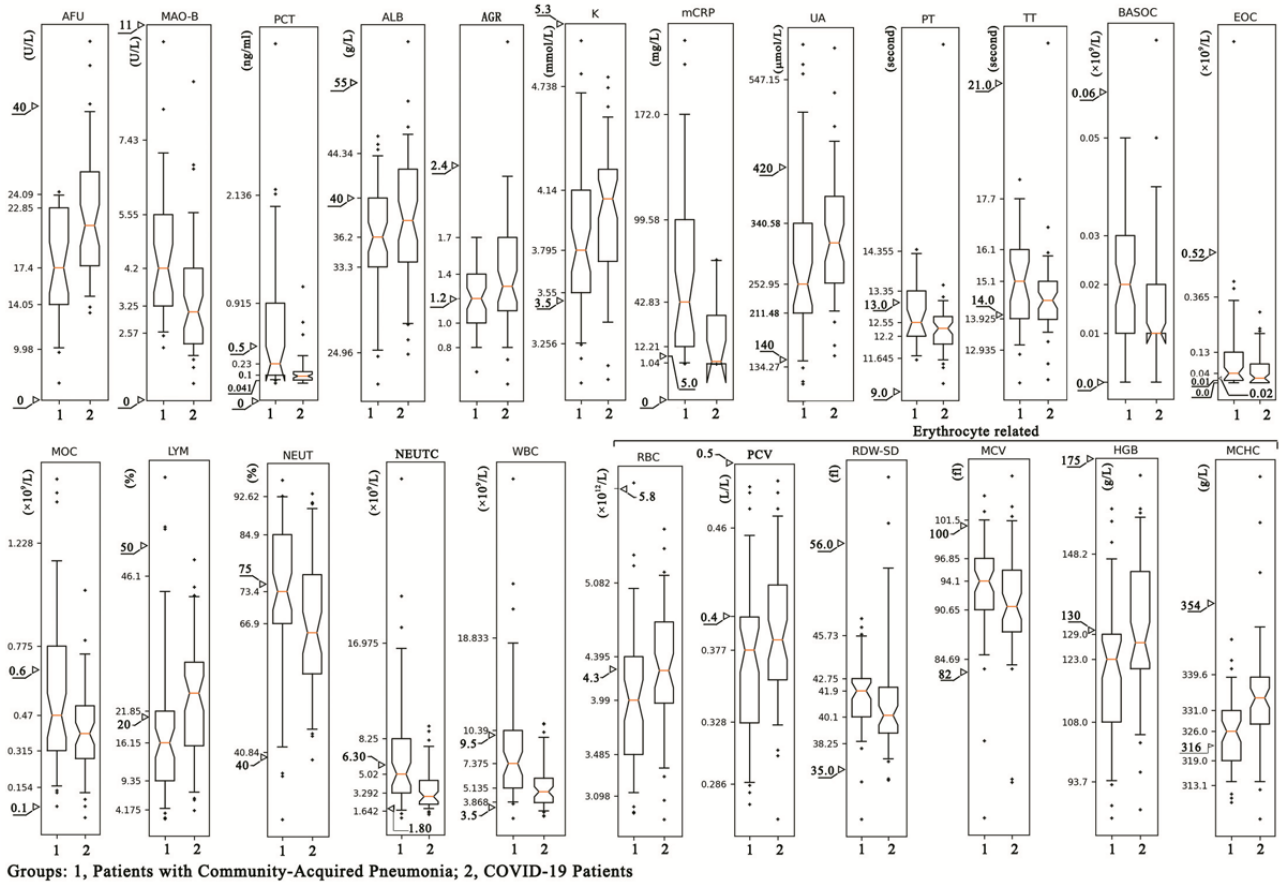
Characteristic Profile of the CLIs in COVID-19 and CAP

Even though most CLIs had a similar variation trend in both CAP and COVID-19, the extent of change was different. Among more than 60 evaluated CLIs, there were significant differences in 25 CLIs between the two groups (see Table 2). A decrease of lymphocyte, red blood cell (RBC) count, hematocrit or packed-cell volume (PCV), hemoglobin concentration, and mean corpuscular hemoglobin concentration (MCHC) and an increase of neutrophil ratio, prothrombin time (PT), micro-CRP (mCRP), and PCT were observed in both COVID-19 and CAP patients. Furthermore, the neutrophil ratio and levels of PT, mCRP, and PCT in CAP were remarkably higher than those in COVID-19. Levels of lymphocyte, RBC count, PCV, hemoglobin concentration, and MCHC in CAP were significantly lower than those in COVID-19 (see Figure 1). Various erythrocyte-related CLIs—RBC count, PCV, hemoglobin concentration, and MCHC—significantly decreased in both CAP and COVID-19, but there was a greater reduction in CAP patients (see Figure 1). The RBC distribution width–standard deviation (RDW-SD) and RBC mean corpuscular volume (MCV) also indicated prominent differences between CAP and COVID-19 (see Figure 1).

Table 2. Differences in clinical laboratory indicators (CLIs) between patients with community-acquired pneumonia (CAP) and COVID-19.

CLI	CAP patients (n=60)		COVID-19 patients (n=61)		*P* value
	n (%)	Mean (SD)	n (%)	Mean (SD)	
Procalcitonin (ng/mL)	43 (72)	0.629 (0.838)	55 (90)	0.134 (0.184)	<.001
Monoamine oxidase B (U/L)	35 (58)	4.569 (1.748)	53 (87)	3.538 (1.592)	.001
Myoglobin (ng/mL)	14 (23)	39.179 (29.421)	23 (38)	65.794 (87.039)	.04
Micro–C-reactive protein (mg/L)	41 (68)	63.943 (64.530)	13 (21)	22.568 (29.577)	.004
Prothrombin time (seconds)	30 (50)	12.780 (0.873)	53 (87)	12.460 (1.107)	.04
Thrombin time (seconds)	30 (50)	15.123 (1.565)	53 (87)	14.655 (1.422)	.049
Albumin (g/L)	53 (88)	35.508 (5.929)	54 (89)	37.831 (6.169)	.04
Albumin to globulin ratio	53 (88)	1.211 (0.295)	54 (89)	1.378 (0.482)	.047
α-L-fucosidase (U/L)	35 (58)	17.709 (5.167)	50 (82)	22.106 (5.698)	<.001
Uric acid (μmol/L)	44 (73)	284.193 (118.608)	54 (89)	325.261 (92.914)	.007
Potassium (mmol/L)	54 (90)	3.900 (0.462)	55 (90)	4.021 (0.392)	.03
White blood count cell ($\times 10^9$/L)	58 (97)	8.858 (5.576)	56 (92)	5.293 (2.047)	<.001
Neutrophils (%)	57 (95)	72.958 (15.544)	56 (92)	66.661 (14.013)	.007
Lymphocytes (%)	56 (93)	18.646 (13.416)	56 (92)	24.014 (11.175)	.002
Neutrophil count ($\times 10^9$/L)	56 (93)	6.797 (5.525)	56 (92)	3.649 (1.949)	<.001
Monocyte count ($\times 10^9$/L)	55 (92)	0.565 (0.337)	56 (92)	0.404 (0.194)	.009
Eosinophil count ($\times 10^9$/L)	55 (92)	0.111 (0.213)	56 (92)	0.053 (0.072)	.03
Basophil count ($\times 10^9$/L)	55 (92)	0.021 (0.013)	56 (92)	0.015 (0.013)	.002
Red blood cell count ($\times 10^{12}$/L)	56 (93)	4.028 (0.647)	56 (92)	4.284 (0.570)	.008
Hemoglobin concentration (g/L)	55 (92)	120.800 (17.326)	56 (92)	130.143 (16.888)	.005
Packed-cell volume (hematocrit) (L/L)	55 (92)	0.371 (0.052)	56 (92)	0.389 (0.049)	.04
Mean red blood cell volume (fL)	55 (92)	93.255 (6.662)	56 (92)	91.241 (6.501)	.01
Mean corpuscular hemoglobin concentration (g/L)	55 (92)	325.473 (8.360)	56 (92)	334.482 (13.559)	<.001
Red blood cell distribution width–standard deviation (fL)	55 (92)	41.476 (2.573)	56 (92)	41.141 (4.082)	.01

Figure 1. The statistical distribution of the plasma level of the clinical laboratory indicators (CLIs) with a remarkable difference between COVID-19 and community-acquired pneumonia (CAP). The statistical distribution was presented with a box and whisker plot. The horizontal lines within the boxes indicate the median value. The vertical lines extending below and above the boxes represent 5%-95% percentile values. The scale on the y-axis represents the values of the 5th, 25th, 50th, 75th, and 95th percentiles of the CLI in the CAP group. The triangles represent the upper and lower limits of the normal reference range of the laboratory index. AFU: α-L-fucosidase; AGR: albumin to globulin ratio; ALB: albumin; BASOC: basophil count; EOC: eosinophil count; HGB: hemoglobin concentration; K: potassium; LYM: lymphocyte; MAO-B: monoaminoxidase B; MCHC: mean corpuscular hemoglobin concentration; mCRP: micro–C-reactive protein; MCV: mean (red blood cell) corpuscular volume; MOC: monocyte count; NEUT: neutrophil ratio; NEUTC: neutrophil count; PCT: procalcitonin; PCV: packed-cell volume (hematocrit); PT: prothrombin time; RBC: red blood cell count; RDW-SD: red blood cell distribution width–standard deviation; TT: thrombin time; UA: uric acid; WBC: white blood cell count.



Comparing the COVID19-COM and COVID19-SV subgroups, 26 CLIs demonstrated a remarkable difference (see Table 3). In comparison with the COVID19-COM subgroup, LDH, aspartate aminotransferase, fibrinogen content, mCRP, and erythrocyte sedimentation rate increased acutely in the COVID19-SV subgroup, whereas prealbumin, carbon dioxide binding capacity, lymphocytes, and lymphocyte count decreased in the COVID19-SV subgroup (see Multimedia Appendix 1).

An orderly increase of α-L-fucosidase (AFU), myoglobin, uric acid, and MCHC and an orderly decrease of thrombin time, monocyte count, eosinophil count, RBC MCV, and RDW-SD were observed in CAP, COVID19-COM, and COVID19-SV patients, indicating that these CLIs may be used to distinguish CAP from COVID-19 and may suggest the probability of severe COVID-19 progression (see Multimedia Appendix 2).

Table 3. Difference in clinical laboratory indicators (CLIs) between patients with common and severe types of COVID-19.

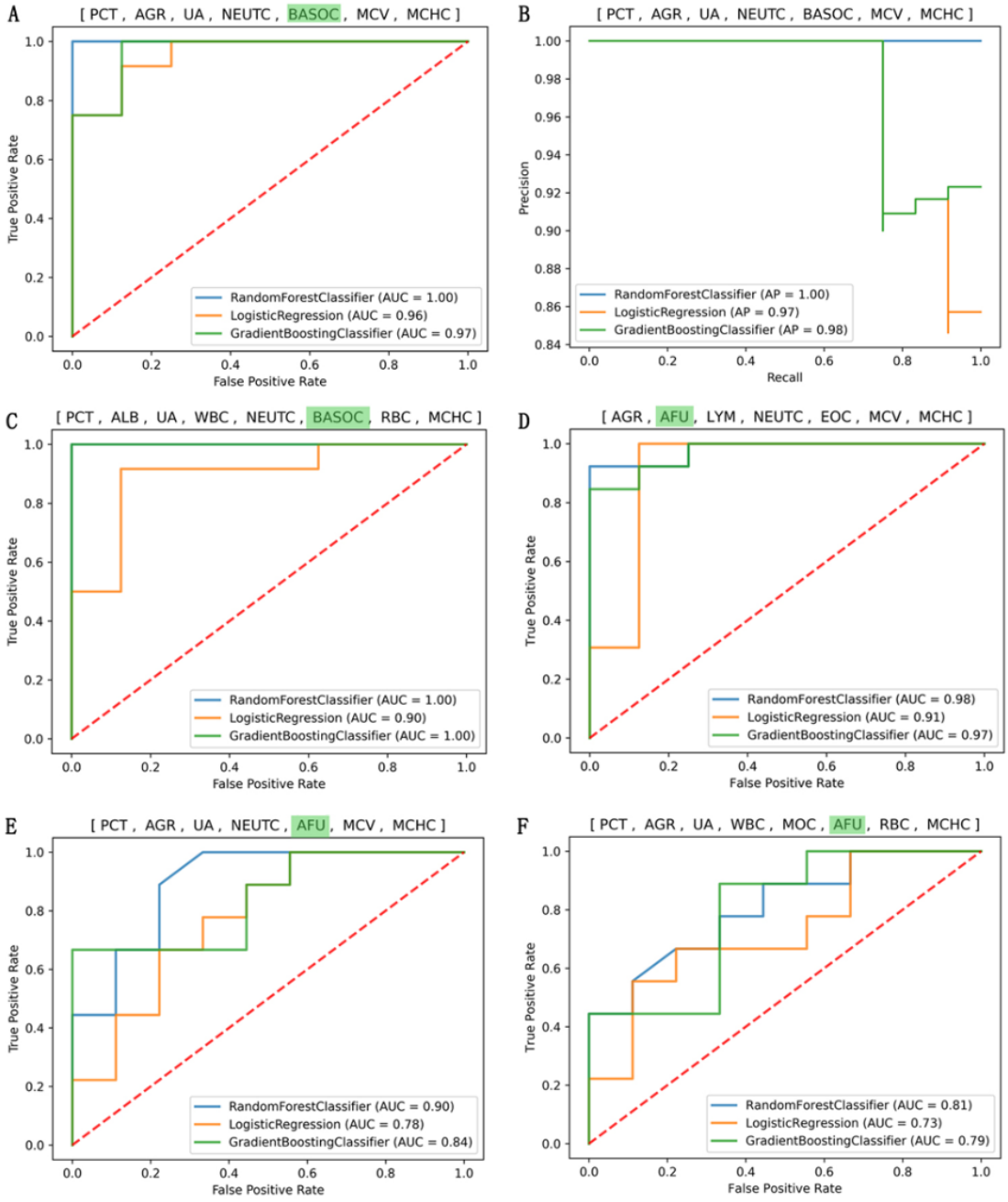
CLIs	Patients with a common type of COVID-19 (n=50)		Patients with a severe type of COVID-19 (n=11)		*P* value
	n (%)	Mean (SD)	n (%)	Mean (SD)	
Procalcitonin (ng/mL)	44 (88)	0.112 (0.170)	11 (100)	0.224 (0.217)	.01
N-terminal pro-B-type natriuretic peptide (pg/mL)	29 (58)	366.053 (549.429)	11 (100)	534.782 (398.067)	.03
Hypersensitive C-reactive protein (mg/L)	41 (82)	23.332 (34.483)	11 (100)	72.458 (60.805)	.002
Lactate dehydrogenase (U/L)	26 (52)	214.896 (73.319)	8 (73)	314.750 (118.755)	.02
D-dimer (mg/L)	42 (84)	0.834 (1.115)	11 (100)	5.133 (10.399)	.005
Myoglobin (ng/mL)	16 (32)	49.221 (60.505)	7 (64)	103.674 (127.354)	.02
Cardiac troponin (ng/mL)	16 (32)	0.011 (0.003)	7 (64)	0.033 (0.041)	.02
Creatine kinase (U/L)	27 (54)	81.296 (47.153)	8 (73)	202.125 (195.052)	.02
Fibrinogen content (mg/dL)	42 (84)	411.905 (104.363)	11 (100)	467.455 (76.500)	.03
Aspartate aminotransferase (U/L)	46 (92)	29.413 (15.756)	10 (91)	45.600 (18.969)	.004
γ-glutamyl transpeptidase (U/L)	44 (88)	46.046 (41.609)	10 (91)	80.000 (44.229)	.007
Albumin (g/L)	44 (88)	38.602 (6.267)	10 (91)	34.440 (4.558)	.02
Albumin to globulin ratio	44 (88)	1.436 (0.507)	10 (91)	1.120 (0.230)	.02
Indirect bilirubin (μmol/L)	44 (88)	9.482 (3.841)	10 (91)	7.960 (4.336)	.048
Prealbumin (mg/L)	41 (82)	180.171 (83.374)	9 (82)	125.556 (68.182)	.03
β2-microglobulin (mg/L)	41 (82)	1.978 (0.430)	9 (82)	2.528 (1.015)	.01
Carbon dioxide binding capacity (mmol/L)	41 (82)	25.420 (2.537)	9 (82)	22.733 (2.018)	.002
Potassium (mmol/L)	44 (88)	4.057 (0.414)	11 (100)	3.876 (0.251)	.04
Erythrocyte sedimentation rate (mm/h)	30 (60)	55.433 (41.639)	7 (64)	87.000 (35.081)	.02
Neutrophils (%)	45 (90)	64.496 (13.286)	11 (100)	75.519 (14.001)	.02
Lymphocytes (%)	45 (90)	25.711 (10.932)	11 (100)	17.073 (9.750)	.01
Eosinophils (%)	45 (90)	1.236 (1.388)	11 (100)	0.391 (1.038)	.009
Eosinophil count ($\times 10^9$/L)	45 (90)	0.062 (0.076)	11 (100)	0.014 (0.039)	.003
Lymphocyte count ($\times 10^9$/L)	45 (90)	1.255 (0.558)	11 (100)	0.835 (0.383)	.008
Packed-cell volume (hematocrit) (L/L)	45 (90)	0.395 (0.050)	11 (100)	0.368 (0.036)	.03
Red blood cell distribution width–coefficient of variation (%)	45 (90)	12.658 (1.171)	11 (100)	12.873 (0.781)	.03

Classifiers Constructed From the FCs With Seven to Eight CLIs Could Accurately Distinguish COVID-19 From CAP

The performance of the classifiers gradually improved as the number of CLIs in the FCs increased from one to eight. However, when the number of CLIs in the FCs reached eight, the performance of the classifiers constructed by these FCs no longer significantly improved. The performance of the LR classifier algorithm constructed with the FCs with eight CLIs (ie, 8-CLI combination) was even slightly lower than those constructed by the FCs with seven CLIs (ie, 7-CLI combination). A total of 43 FCs, including five 7-CLI combinations and 38 8-CLI combinations, were determined according to the recall rate. The AUROCs of the classifiers constructed with the LR classifier, RFC, and GBC algorithms were greater than 0.85 (see Multimedia Appendix 3, Table S1). The AUROC and precision-recall curves of the classifiers constructed with the RFC, LR classifier, and GBC algorithms from the representative 7-CLI combination (ie, PCT, albumin to globulin ratio [AGR], uric acid, neutrophil count, basophil count, RBC MCV, and MCHC) showed very high performance and precision in COVID-19 prediction; their AUROCs were 1.0, 0.97, and 0.96, respectively (see Figure 2, A), and their average precision values were 1.0, 0.97, and 0.98, respectively (Figure 2, B). The AUROCs of the classifiers constructed with the RFC, LR classifier, and GBC algorithms from the representative 8-CLI combination (ie, PCT, albumin, uric acid, WBC [white blood cell] count, monocyte count, basophil count, RBC count, and MCHC) were 1.0, 0.90, and 1.0, respectively (see Figure 2, C). The AUROCs of the classifiers constructed with the three algorithms from the 7-CLI combination (ie, agr, afu, lymphocytes, neutrophil counts, eosinophil count, RBC mcv, and mchc) were 0.98, 0.91, and 0.97, respectively (see Figure

2, D). Feature importance results showed that basophil count was the least important in the above two representative CLI combinations, and AFU was the most important in the CLI combinations (see Figure 3). However, when basophil count was substituted with AFU in the two above-mentioned CLI combinations, the performance of the classifiers constructed with the new CLI combinations decreased (see Figure 2, E and F). PCT and AFU were not observed to be in the same CLI combination from which an HPC could be constructed. The evidence above and the fact that only 43 FCs with seven or eight CLIs could be used to build HPCs suggested that only the FCs with specific CLIs can establish HPCs to distinguish COVID-19 from CAP.

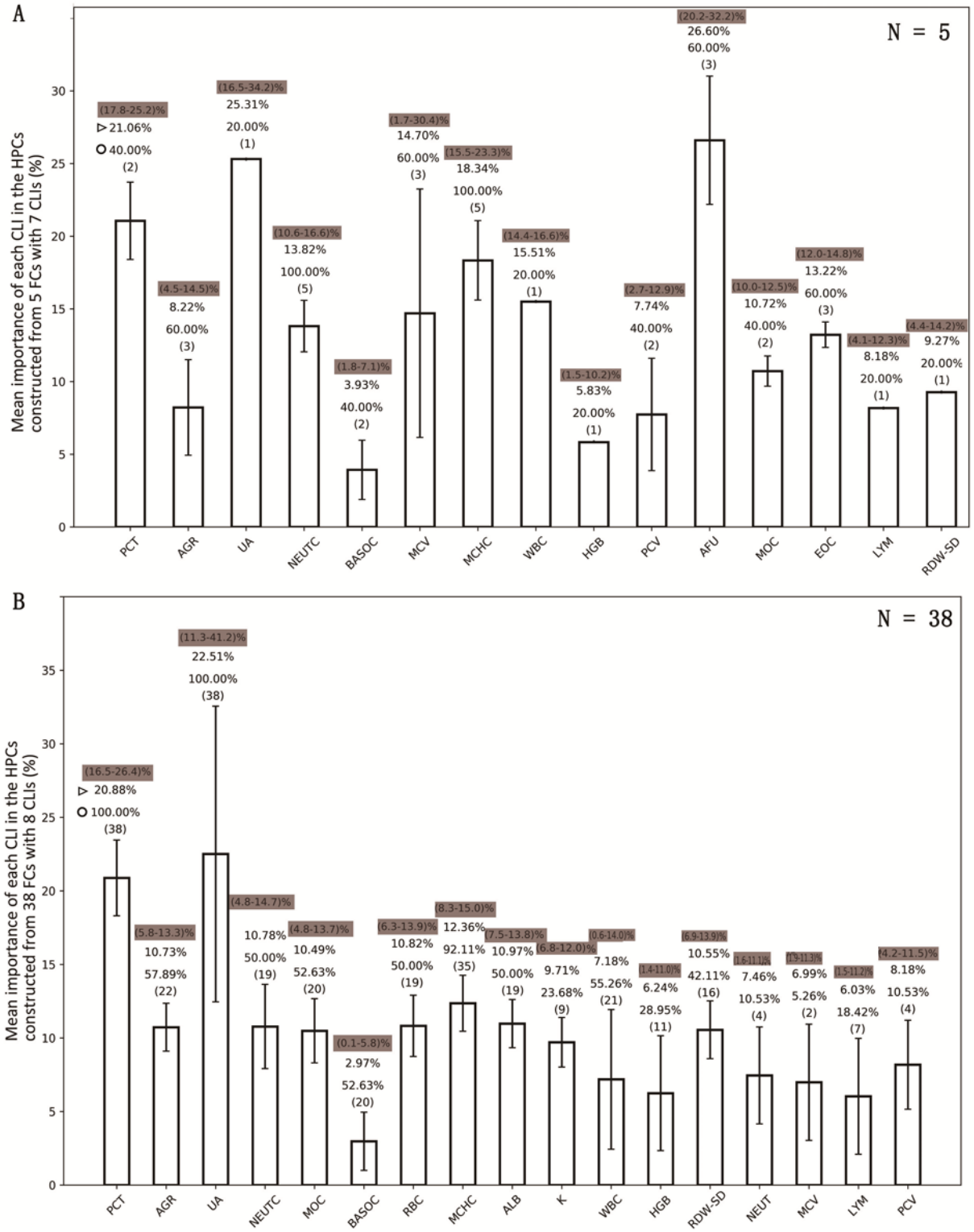
Figure 2. Area under the receiver operating characteristic curve (AUROC) and precision-recall curve plotted for the COVID-19 vs community-acquired pneumonia (CAP) classifiers built with various feature combinations (FCs) of different clinical laboratory indicators (CLIs). At the top of each image is the CLI combination for constructing classifiers using three different classification algorithms. AFU: α-L-fucosidase; AGR: albumin to globulin ratio; ALB: albumin; BASOC: basophil count; EOC: eosinophil count; LYM: lymphocyte; MCHC: mean corpuscular hemoglobin concentration; MCV: mean (red blood cell) corpuscular volume; MOC: monocyte count; NEUTC: neutrophil count; PCT: procalcitonin; RBC: red blood cell count; UA: uric acid; WBC: white blood cell count.



The importance of different CLIs in classifiers varied greatly, and the importance of the same CLI varied greatly among classifiers constructed by different FCs (see Figure 3). In the HPCs constructed with the 7-CLI combinations, the average feature importance of AFU (26.60%) was the highest, followed by uric acid (25.31%) and PCT (21.06%) (see Figure 3, A). However, in the HPCs constructed with the 8-CLI combinations, the average feature importance of uric acid (22.51%) was the

highest, followed by PCT (20.88%) and MCHC (12.36%) (see Figure 3, B). PCT and MCHC were very important to each classifier because they were included, respectively, in 100% (38/38) and 92% (35/38) of the 8-CLI combinations (see Figure 3, B) and in 40% (2/5) and 100% (5/5) of the 7-CLI combinations (see Figure 3, A). Uric acid was also included in all 8-CLI combinations, but its feature importance varied from 11.3% to 41.2% in different classifiers (see Figure 3, B).

Figure 3. Usage rate and the feature importance of each clinical laboratory indicator (CLI) in the high-performance COVID-19 vs community-acquired pneumonia (CAP) classifiers. (A) The mean feature importance of each CLI in the high-performance classifiers (HPCs) constructed with the 7-CLI combinations. (B) The mean feature importance of each CLI in the HPCs constructed with the 8-CLI combinations. The histogram is represented by mean (SD). The numbers with the shadow backgrounds represent the minimum and maximum values of the feature importance of the CLI. The number indicated with the triangle symbol represents the mean feature importance of CLI in all classifiers. The number indicated with the circle represents the usage rate of the CLI in the HPC. The number in the parentheses indicates how many CLI combinations are capable of constructing the HPCs containing the CLI. AFU: α-L-fucosidase; AGR: albumin to globulin ratio; ALB: albumin; BASOC: basophil count; EOC: eosinophil count; FC: feature combination; HGB: hemoglobin concentration; K: potassium; LYM: lymphocyte; MCHC: mean corpuscular hemoglobin concentration; MCV: mean (red blood cell) corpuscular volume; MOC: monocyte count; NEUT: neutrophil ratio; NEUTC: neutrophil count; PCT: procalcitonin; PCV: packed-cell volume (hematocrit); RBC: red blood cell count; RDW-SD: red blood cell distribution width–standard deviation; UA: uric acid; WBC: white blood cell count.



Discussion

Principal Findings

The main highlight of this study is that only a few of the common CLIs were required to establish the classifier models to accurately distinguish COVID-19 from CAP. The HPCs could only be constructed by combining several specific CLIs. Among the nearly 2 million FCs with one to eight CLIs, only 43 FCs could be used to construct HPCs with a recall rate greater than 0.9 and an AUROC greater than 0.85 to distinguish COVID-19 from CAP.

Comparison With Prior Work

We have established many COVID-19 versus CAP HPCs with FCs consisting of only CLIs, and almost no similar research results on distinguishing COVID-19 from CAP have been reported. However, many studies have used CLIs to build ML models to help with COVID-19 diagnosis. The prediction performance of these models varied: the accuracy of these models in predicting COVID-19 was between 0.8 and 0.96 [30-32]. In addition, most of the reported ML models for the diagnosis or prediction of COVID-19 have involved more types of variables, such as CT results, clinical symptoms, and CLIs [17,32,33]. Although most of these COVID-19-related ML models were built with more than two ML algorithms, not all models constructed with each algorithm showed high performance. The methods of feature selection that were used in these studies included the recursive feature elimination algorithm [31], causal explanation models [17], and the least absolute shrinkage and selection operator regression [32]. These methods can extract the features that are closely related to the target phenotype, but whether the classifier constructed by the combination of these features has the best performance needs to be determined. The optimized FCs in this study were selected by evaluating the recall rate and AUROC for each FC with one to eight randomly selected CLIs from the differential CLIs between COVID-19 and CAP groups and by constructing classifiers using each FC with the LR classifier algorithm. The FCs that were preliminarily screened were used to build classifiers with RFC and GBC algorithms; finally, only the FCs capable of building the HPC simultaneously with the LR classifier, RFC, and GBC algorithms were selected for the final model construction.

Limitations

As reported earlier, many inflammatory factors, including IL-6 and interleukin-10 (IL-10), are closely related to COVID-19 and have diagnostic value, but neither IL-6 nor IL-10 were detected in the patients of this study. Menni et al [18] reported that loss of smell and taste is a strong predictor for COVID-19. Deviations and omissions may exist in the patients' self-reported clinical symptoms. Thus, we did not take into account the clinical symptoms when building the classifiers. The possibility that other indicators are more important in constructing COVID-19 versus CAP classifiers was not ruled out. In addition, the sample size included in this study was relatively small, and the classifiers need to be optimized with larger samples before it can be used to distinguish COVID-19 from CAP in practice.

The Rationality of the Research Results

Out of the 43 FCs, 40 contained PCT and MCHC. The feature importance of PCT in each classifier is very high, suggesting that PCT may be a good blood marker to efficiently distinguish COVID-19 from CAP. PCT is one of the markers of lower respiratory tract bacteria and other infections. The US Food and Drug Administration approved the monitoring of the beginning and the entire duration of antibiotic treatment for suspected lower respiratory tract infections based on serum PCT levels [12]. However, the elevation of serum PCT in COVID-19 patients was also reported in many studies [34]. The increase of PCT is a remarkable characteristic of patients with COVID-19 [34]. Increased serum PCT levels in both COVID-19 and CAP patients indicated that the distinction of COVID-19 from CAP could not be made simply on the basis of the increase in PCT. Compared with the normal reference values of the CLIs, the serum levels of most of the CLIs increased or decreased simultaneously in both COVID-19 and CAP patients. Thus, providing references for the diagnosis of COVID-19 or CAP directly in regard to the rise or decrease of the CLIs is difficult. However, we found that the ML classifiers constructed with the FCs with many certain CLIs could distinguish COVID-19 from CAP effectively, suggesting an advantage of ML algorithms in disease classification or diagnosis.

The COVID-19 versus CAP classifiers with the highest performance also involved PCT, MCHC, uric acid, albumin, neutrophil count, monocyte count, basophil count, RBC count, and WBC count, proposing the importance of these CLIs in differentiating COVID-19 from CAP. Few studies have reported the changing trend of MCHC in patients with COVID-19 or CAP, but the results of this study showed that MCHC decreased in both groups and was significantly lower in the CAP group than in the COVID-19 group. The reason for the decrease of MCHC may be closely related to the reduction of iron due to inflammation [35]. The IQRs of uric acid in both COVID-19 and CAP groups were within the normal reference range, but the IQR was significantly higher in the COVID-19 group than in the CAP group. Elevated uric acid is an independent risk factor of renal injury or renal dysfunction; the underlying mechanisms of uric acid elevation are very complicated [36]. The significant difference in uric acid between COVID-19 and CAP may be interpreted as follows: individuals with higher uric acid may be more susceptible to COVID-19 than those with lower uric acid levels. Uric acid exists in all 8-CLI combinations that are capable of constructing high-performance CLIs and has a high feature importance in the classifiers, suggesting that uric acid is another important marker that can distinguish COVID-19 from CAP. Zhou et al reported that albumin significantly decreased in severe and critical COVID-19 patients [37]. Serum albumin level is a good prognostic marker in CAP. A decreased albumin level is closely associated with a higher risk of mortality in patients with CAP [38]. Although albumin decreased remarkably in both COVID-19 and CAP groups, there was still a significant difference between the two groups; the decrease in the CAP group was more obvious than that in the COVID-19 group, which could contribute to the differentiation of COVID-19 from CAP. AFU contributed high feature importance in the HPCs constructed from 7-CLI combinations

due to the significant difference in AFU between COVID-19 and CAP. An increase of serum AFU has a certain diagnostic value for primary liver cancer [39]. Thus, the higher AFU in the COVID-19 group than in the CAP group may be explained by the fact that liver injury is more common in COVID-19 than in CAP or that the diversity in AFU levels determines the difference in susceptibility to COVID-19.

Recommendations

Both PCT and AFU contributed high feature importance in the HPCs constructed from the FCs containing PCT or AFU, but the performance of the classifiers constructed from the FCs containing both PCT and AFU decreased remarkably. This result indicated that intrinsic dependence exists among some CLIs that undergo synergistic changes in individuals and can be used to construct HPCs. The internal relationship between CLIs is very complex and difficult to deconstruct. Therefore, the following method may be effective: random selection of different CLIs to construct classifiers with different classification algorithms, followed by the evaluation of the performance of each classifier, and, finally, the discovery of the FCs with certain CLIs that can be used to accurately distinguish COVID-19 from CAP.

Conclusions

The patients suffering from COVID-19 and CAP have their own characteristic profiles of CLIs, and some FCs consisting of seven or eight specific CLIs could build COVID-19 versus CAP HPCs. The usage rate and the feature importance of the CLIs in the HPCs indicated that PCT, MCHC, uric acid, albumin, AGR, neutrophil count, RBC count, monocyte count, and WBC count are the most important indicators that can distinguish COVID-19 from CAP.

Acknowledgments

We are very grateful to Associate Professor Benrong Liu of the Second Affiliated Hospital of Guangzhou Medical University for his guidance in data analysis. We are also very grateful to the doctors and nurses of Gong An County People's Hospital for their hard work to prevent as well as control COVID-19.

Authors' Contributions

XZH (huangxz020@gzucm.edu.cn) and YX (yujuanxiong@gzucm.edu.cn) share corresponding author duties for this article. XZH and YX made substantial contributions to the study concept and design. YX and PFK were in charge of drafting the manuscript. WD and WH were responsible for obtaining ethical approval, collecting data, and confirming the accuracy of the data. ZZL, QZZ, and YW performed the data analysis and interpretation. All authors agreed on the final version for submission to the journal.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The statistical distribution of the plasma level of the clinical laboratory indicators (CLIs) with a significant difference between COVID19-COM (COVID-19 patient subgroup with mild and common types) and COVID19-SV (COVID-19 patient subgroup with severe and critical types). The statistical distribution was presented with a box and whisker plot. The horizontal lines within the boxes indicate the median value. The vertical lines extending below and above the boxes represent 5%-95% percentile values. The scale on the y-axis represents the 5th, 25th, 50th, 75th, and 95th percentile values of the CLI in the COVID19-COM subgroup. The triangles represent the upper and lower limits of the normal reference range of the laboratory index. The median of the CLI in the COVID19-SV subgroup is also represented in the y-axis. AST: aspartate aminotransferase; CO2CP: carbon dioxide binding capacity; ESR: erythrocyte sedimentation rate; γ-GGT: transglutaminase transpeptidase gamma; FIB: fibrinogen content; LDH: lactate dehydrogenase; LYM: lymphocyte; LYMPH: lymphocyte count; mCRP: micro–C-reactive protein; MYO: myoglobin; NEUT: neutrophil ratio; PA: prealbumin.
[PNG File , 403 KB - jmir_v23i2e23390_app1.png]

Multimedia Appendix 2

The statistical distribution of the plasma level of the clinical laboratory indicators (CLIs) among community-acquired pneumonia (CAP), COVID19-COM (COVID-19 patient subgroup with mild and common types), and COVID19-SV (COVID-19 patient subgroup with severe and critical types). The statistical distribution was presented with a box and whisker plot. The horizontal lines within the boxes indicate the median value. The vertical lines extending below and above the boxes represent 5%-95% percentile values. The scale on the y-axis represents the 5th, 25th, 50th, 75th, and 95th percentile values of the CLI in the CAP group. The triangles represent the upper and lower limits of the normal reference range of the laboratory index. AFU: α-L-fucosidase; EOC: eosinophil count; MCHC: mean corpuscular hemoglobin concentration; MCV: mean (red blood cell) corpuscular volume; MOC: monocyte count; MYO: myoglobin; RDW-SD: red blood cell distribution width–standard deviation; TT: thrombin time; UA: uric acid.

[PNG File , 204 KB - jmir_v23i2e23390_app2.png]

Multimedia Appendix 3
Clinical laboratory indicator (CLI) combinations and the hyper-parameters of the classifiers constructed by different machine learning algorithms from these CLI combinations.
[XLSX File (Microsoft Excel File), 14 KB - jmir_v23i2e23390_app3.xlsx]

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-05-06]
2. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020 Apr;5(4):536-544 [FREE Full text] [doi: 10.1038/s41564-020-0695-z] [Medline: 32123347]
3. Ye F, Xu S, Rong Z, Xu R, Liu X, Deng P, et al. Delivery of infection from asymptomatic carriers of COVID-19 in a familial cluster. Int J Infect Dis 2020 May;94:133-138 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.042] [Medline: 32247826]
4. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
5. Zhang J, Tian S, Lou J, Chen Y. Familial cluster of COVID-19 infection from an asymptomatic. Crit Care 2020 Mar 27;24(1):119 [FREE Full text] [doi: 10.1186/s13054-020-2817-7] [Medline: 32220236]
6. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720. [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
7. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N Engl J Med 2020 Mar 26;382(13):1199-1207 [FREE Full text] [doi: 10.1056/NEJMoa2001316] [Medline: 31995857]
8. Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung S, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. J Clin Med 2020 Feb 17;9(2):538 [FREE Full text] [doi: 10.3390/jcm9020538] [Medline: 32079150]
9. Terpos E, Ntanasis-Stathopoulos I, Elalamy I, Kastritis E, Sergentanis TN, Politou M, et al. Hematological findings and complications of COVID-19. Am J Hematol 2020 Jul;95(7):834-847 [FREE Full text] [doi: 10.1002/ajh.25829] [Medline: 32282949]
10. Ji D, Zhang D, Xu J, Chen Z, Yang T, Zhao P, et al. Prediction for progression risk in patients with COVID-19 pneumonia: The CALL score. Clin Infect Dis 2020 Sep 12;71(6):1393-1399 [FREE Full text] [doi: 10.1093/cid/ciaa414] [Medline: 32271369]
11. Wang K, Zuo P, Liu Y, Zhang M, Zhao X, Xie S, et al. Clinical and laboratory predictors of in-hospital mortality in patients with coronavirus disease-2019: A cohort study in Wuhan, China. Clin Infect Dis 2020 Nov 19;71(16):2079-2088 [FREE Full text] [doi: 10.1093/cid/ciaa538] [Medline: 32361723]
12. Kamat IS, Ramachandran V, Eswaran H, Guffey D, Musher DM. Procalcitonin to distinguish viral from bacterial pneumonia: A systematic review and meta-analysis. Clin Infect Dis 2020 Jan 16;70(3):538-542. [doi: 10.1093/cid/ciz545] [Medline: 31241140]
13. Robriquet L, Guery B, Fourrier F. Pulmonary coagulopathy as a new target in therapeutic studies of acute lung injury or pneumonia. Crit Care Med 2006 Sep;34(9):2510; author reply 2510-2510; author reply 2511. [doi: 10.1097/01.CCM.0000235673.40874.BD] [Medline: 16921333]
14. Schultz MJ, Haitsma JJ, Zhang H, Slutsky AS. Pulmonary coagulopathy as a new target in therapeutic studies of acute lung injury or pneumonia--A review. Crit Care Med 2006 Mar;34(3):871-877. [Medline: 16521285]
15. Ceccato A, Panagiotarakou M, Ranzani OT, Martin-Fernandez M, Almansa-Mora R, Gabarrus A, et al. Lymphocytopenia as a predictor of mortality in patients with ICU-acquired pneumonia. J Clin Med 2019 Jun 13;8(6):843 [FREE Full text] [doi: 10.3390/jcm8060843] [Medline: 31200458]
16. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: Machine learning and the future of medicine. J Intern Med 2018 Dec;284(6):603-619 [FREE Full text] [doi: 10.1111/joim.12822] [Medline: 30102808]
17. Schwab P, DuMont Schütte A, Dietz B, Bauer S. Clinical predictive models for COVID-19: Systematic study. J Med Internet Res 2020 Oct 06;22(10):e21439 [FREE Full text] [doi: 10.2196/21439] [Medline: 32976111]
18. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]

19. Ko J, Baldassano SN, Loh P, Kording K, Litt B, Issadore D. Machine learning to detect signatures of disease in liquid biopsies - A user's guide. Lab Chip 2018 Jan 30;18(3):395-405 [FREE Full text] [doi: 10.1039/c7lc00955k] [Medline: 29192299]
20. Chinese Clinical Guidance for COVID-19 Pneumonia Diagnosis and Treatment (7th edition). Beijing, China: China National Health Commission; 2020 Mar 04. URL: http://kjfy.meetingchina.org/msite/news/show/cn/3337.html [accessed 2020-05-11]
21. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with Python. In: Proceedings of the 9th Python in Science Conference (SciPy 2010). 2010 Jul 03 Presented at: The 9th Python in Science Conference (SciPy 2010); June 28-July 3, 2010; Austin, TX p. 92-96 URL: https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf [doi: 10.25080/majora-92bf1922-011]
22. itertools — Functions creating iterators for efficient looping. The University of Auckland School of Computer Science. URL: https://www.cs.auckland.ac.nz/references/python/3.2.3-docs/library/itertools.html [accessed 2020-03-08]
23. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
24. Logistic regression, scikit-learn 0.24.1. scikit-learn. URL: https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression [accessed 2020-06-12]
25. Forests of randomized trees, scikit-learn 0.24.1. scikit-learn. URL: https://scikit-learn.org/stable/modules/ensemble.html#forests-of-randomized-trees [accessed 2020-06-12]
26. Breiman L. Random forests. Mach Learn 2001 Oct;45:5-32 [FREE Full text] [doi: 10.1023/A:1010933404324]
27. Gradient tree boosting, scikit-learn 0.24.1. scikit-learn. URL: https://scikit-learn.org/stable/modules/ensemble.html#gradient-boosting [accessed 2020-06-12]
28. Friedman JH. 1999 Reitz Lecture. Greedy function approximation: A gradient boosting machine. Ann Stat 2001;29(5):1189-1232 [FREE Full text]
29. feature_importances_, scikit-learn 0.24.1. scikit-learn. URL: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn.ensemble.RandomForestClassifier.feature_importances_ [accessed 2021-02-12]
30. McDonald SA, Medford RJ, Basit MA, Diercks DB, Courtney DM. Derivation with internal validation of a multivariable predictive model to predict COVID-19 test results in emergency department patients. Acad Emerg Med 2020 Nov 28:1-9 [FREE Full text] [doi: 10.1111/acem.14182] [Medline: 33249683]
31. Yao H, Zhang N, Zhang R, Duan M, Xie T, Pan J, et al. Severity detection for the coronavirus disease 2019 (COVID-19) patients using a machine learning model based on the blood and urine tests. Front Cell Dev Biol 2020;8:683 [FREE Full text] [doi: 10.3389/fcell.2020.00683] [Medline: 32850809]
32. Gong J, Ou J, Qiu X, Jie Y, Chen Y, Yuan L, et al. A tool for early prediction of severe coronavirus disease 2019 (COVID-19): A multicenter study using the risk nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020 Jul 28;71(15):833-840 [FREE Full text] [doi: 10.1093/cid/ciaa443] [Medline: 32296824]
33. Schwab P, DuMont Schütte A, Dietz B, Bauer S. Clinical predictive models for COVID-19: Systematic study. J Med Internet Res 2020 Oct 06;22(10):e21439 [FREE Full text] [doi: 10.2196/21439] [Medline: 32976111]
34. Zhu Z, Cai T, Fan L, Lou K, Hua X, Huang Z, et al. Clinical value of immune-inflammatory parameters to assess the severity of coronavirus disease 2019. Int J Infect Dis 2020 Jun;95:332-339 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.041] [Medline: 32334118]
35. Huang Y, Hu Z. Lower mean corpuscular hemoglobin concentration is associated with poorer outcomes in intensive care unit admitted patients with acute myocardial infarction. Ann Transl Med 2016 May;4(10):190 [FREE Full text] [doi: 10.21037/atm.2016.03.42] [Medline: 27294086]
36. Cao X, Wu L, Chen Z. The association between elevated serum uric acid level and an increased risk of renal function decline in a health checkup cohort in China. Int Urol Nephrol 2018 Mar;50(3):517-525. [doi: 10.1007/s11255-017-1732-6] [Medline: 29094330]
37. Zhou Y, Guo S, He Y, Zuo Q, Liu D, Xiao M, et al. COVID-19 is distinct from SARS-CoV-2-negative community-acquired pneumonia. Front Cell Infect Microbiol 2020;10:322 [FREE Full text] [doi: 10.3389/fcimb.2020.00322] [Medline: 32612961]
38. Viasus D, Garcia-Vidal C, Simonetti A, Manresa F, Dorca J, Gudiol F, et al. Prognostic value of serum albumin levels in hospitalized adults with community-acquired pneumonia. J Infect 2013 May;66(5):415-423. [doi: 10.1016/j.jinf.2012.12.007] [Medline: 23286966]
39. Deugnier Y, David V, Brissot P, Mabo P, Delamaire D, Messner M, et al. Serum alpha-L-fucosidase: A new marker for the diagnosis of primary hepatic carcinoma? Hepatology 1984;4(5):889-892. [doi: 10.1002/hep.1840040516] [Medline: 6207091]

Abbreviations

AFU: α-L-fucosidase
AGR: albumin to globulin ratio
AUROC: area under the receiver operating characteristic curve

CAP: community-acquired pneumonia
CLI: clinical laboratory indicator
COVID19-COM: COVID-19 patient subgroup with mild and common types
COVID19-SV: COVID-19 patient subgroup with severe and critical types
CRP: C-reactive protein
CT: computed tomography
FC: feature combination
GBC: gradient boosting classifier
HPC: high-performance classifier
IL-6: interleukin-6
IL-10: interleukin-10
LDH: lactate dehydrogenase
LR: logistic regression
MCHC: mean corpuscular hemoglobin concentration
mCRP: micro–C-reactive protein
MCV: mean corpuscular volume
ML: machine learning
PCT: procalcitonin
PCV: packed-cell volume
PT: prothrombin time
RBC: red blood cell
RDW-SD: red blood cell distribution width–standard deviation
RFC: random forest classifier
WBC: white blood cell

Edited by G Fagherazzi; submitted 11.08.20; peer-reviewed by J Pacheco, M Elbattah; comments to author 09.12.20; revised version received 29.12.20; accepted 01.02.21; published 22.02.21.

Please cite as:
Dai W, Ke PF, Li ZZ, Zhuang QZ, Huang W, Wang Y, Xiong Y, Huang XZ
Establishing Classifiers With Clinical Laboratory Indicators to Distinguish COVID-19 From Community-Acquired Pneumonia: Retrospective Cohort Study
J Med Internet Res 2021;23(2):e23390
URL: https://www.jmir.org/2021/2/e23390
doi:10.2196/23390
PMID:33534722

Original Paper

Digital Pathology During the COVID-19 Outbreak in Italy: Survey Study

Simone Giaretto[1*], MD; Salvatore Lorenzo Renne[1,2*], MD; Daoud Rahal[2], MD; Paola Bossi[2], MD; Piergiuseppe Colombo[2], MD; Paola Spaggiari[2], MD; Sofia Manara[2], MD; Mauro Sollai[2], MD; Barbara Fiamengo[2], MD; Tatiana Brambilla[2], MD; Bethania Fernandes[2], MD; Stefania Rao[2], MD; Abubaker Elamin[1], MD; Marina Valeri[1,2], MD; Camilla De Carlo[1,2]; Vincenzo Belsito[1,2], MD; Cesare Lancellotti[1,2], MD; Miriam Cieri[2], MD; Angelo Cagini[2], MD; Luigi Terracciano[1,2], MD; Massimo Roncalli[1,2], MD, PhD; Luca Di Tommaso[1,2], MD

[1]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele (MI), Italy

[2]Department of Pathology, Humanitas Clinical and Research Center – IRCCS, Rozzano (MI), Italy

[*]these authors contributed equally

Corresponding Author:
Salvatore Lorenzo Renne, MD
Department of Pathology
Humanitas Clinical and Research Center – IRCCS
via Manzoni 56
Rozzano (MI), 20089
Italy
Phone: 39 0282244787
Email: salvatore.renne@hunimed.eu

Abstract

Background: Transition to digital pathology usually takes months or years to be completed. We were familiarizing ourselves with digital pathology solutions at the time when the COVID-19 outbreak forced us to embark on an abrupt transition to digital pathology.

Objective: The aim of this study was to quantitatively describe how the abrupt transition to digital pathology might affect the quality of diagnoses, model possible causes by probabilistic modeling, and qualitatively gauge the perception of this abrupt transition.

Methods: A total of 17 pathologists and residents participated in this study; these participants reviewed 25 additional test cases from the archives and completed a final psychologic survey. For each case, participants performed several different diagnostic tasks, and their results were recorded and compared with the original diagnoses performed using the gold standard method (ie, conventional microscopy). We performed Bayesian data analysis with probabilistic modeling.

Results: The overall analysis, comprising 1345 different items, resulted in a 9% (117/1345) error rate in using digital slides. The task of differentiating a neoplastic process from a nonneoplastic one accounted for an error rate of 10.7% (42/392), whereas the distinction of a malignant process from a benign one accounted for an error rate of 4.2% (11/258). Apart from residents, senior pathologists generated most discrepancies (7.9%, 13/164). Our model showed that these differences among career levels persisted even after adjusting for other factors.

Conclusions: Our findings are in line with previous findings, emphasizing that the duration of transition (ie, lengthy or abrupt) might not influence the diagnostic performance. Moreover, our findings highlight that senior pathologists may be limited by a digital gap, which may negatively affect their performance with digital pathology. These results can guide the process of digital transition in the field of pathology.

(J Med Internet Res 2021;23(2):e24266) doi:10.2196/24266

KEYWORDS

COVID19; digital pathology; Bayesian data analysis; probabilistic modeling

Introduction

Digital pathology (DP) intends to use computer workstations and digital whole slide imaging to diagnose a pathological process [1-4]. A complete transition from classical to digital pathology is usually a "soft" procedure, taking months or even years to be completed [4-9]. We planned a digitalization of our department, and we were testing several technical aspects related to digital transition. By February 2020, most of our staff pathologists and residents had used digital whole slide imaging for educational or scientific purposes, but the situation radically changed in March 2020. With the COVID-19 pandemic and the subsequent guidelines adopted by the Italian national government and the medical direction of our hospital, we were forced to reduce the presence of staff in the laboratory. Taking advantage of the ongoing digitalization, we decided to adopt DP to sustain smart work.

Most of the reported discordances between diagnoses in DP and those by the gold standard (ie, evaluation of a glass slide under a microscope) are less than 10% [10], and none of these reports were made under an abrupt transition in diagnostic approach. These discrepancies could be attributed to several factors that could be pathologist dependent (eg, career level or individual performance) or pathologist independent (eg, specimen type or the task to be undertaken during the diagnostic procedure). Discerning the relative effect of these features (that could be really small)—even in a carefully designed experimental setting—might be challenging. Probabilistic modeling (and Bayesian data analysis, in general) allows the detection of small effects [11-13]. Moreover, the employment of multilevel hierarchical modeling permits the transfer of shared information among data clusters, resulting in balanced regularization; thus, it reduces overfitting and improves the out-of-sample predictive performance [11,14-18].

In this study, we aimed to (1) quantitatively describe how abrupt transition to DP might affect the quality of diagnosis, (2) model the possible causes via probabilistic modeling, and (3) qualitatively gauge the perception of this abrupt transition.

Methods

A detailed description of the study methods is described in Multimedia Appendix 1 [15,16,19-24].

Ethics Approval

No ethics approval was required for this study. The study participants (ie, pathologists and residents) agreed to—and coauthored—the study.

Study Participants

This study involved 17 participants who were divided into the following 4 groups or career levels, based on their pathology experience: (1) senior (pathologists with >20 years of experience, n=2), (2) expert (pathologists with 10-20 years of experience, n=5), (3) junior (pathologists with <10 years of experience, n=6), and (4) resident (1st year, n=1; 2nd year, n=3). Each of the 17 participants evaluated 25 digital cases, with a total of 425 digital images examined in the study. Overall, 1445 questions were examined (ie, 85 questions per participant) in the study.

Study Design

In addition to their own diagnostic tasks, which were not considered in this study, the pathologists and residents received (1) a set of digital cases within the area of general surgical pathology, (2) specific questions to be addressed while reviewing the cases, and (3) a survey about their digital experience.

Sets of Digital Cases

We set up 5 sets of digital cases representing 3 different specialties (breast: n=2; urology: n=1; and gastrointestinal: n=2) and assigned them to each study participant. Each test comprised 5 cases, represented by one or more slides of a single case that was previously diagnosed using conventional microscopy by the referral pathologist at our institution. The information reported about the original diagnosis was considered as the gold standard. To cover a spectrum of conditions overlapping the routine situation, we considered biopsy and surgical specimens (specimen type). Cases were digitalized using the Aperio AT2 scanner (Leica Biosystems) and visualized using the WebViewer APERIO ImageScope (version 12.1). The slides used for the tests were from 8 nontumoral and 17 tumoral cases. Of the tumoral cases, 7 tumors were benign and 10 were malignant; all malignant tumors were infiltrative and equally distributed between grade 2 and grade 3; 14 cases were biopsy and 11 were surgical.

Study Questionnaire

Participants answered (all or some) of the following questions (ie, categories of diagnostic task), for each case: (1) Is it neoplastic or negative for neoplasia? (2) Is it a malignant (in situ or infiltrative) or a benign neoplasia? (3) What is the histopathological diagnosis? (4) What is the histotype of the lesion? (5) What is the grade of the lesion? Questions 1 and 3 were answered for all cases, question 2 was answered only for neoplastic lesions, and questions 4 and 5 were answered for malignant neoplasms.

Statistical Analysis

To model data clusters, we used a varying effects, multilevel (hierarchical) model [14-16]. The rate of wrong answers (W_i) was modeled as a Bernoulli distribution:

$$W_i \sim Binomial\ (\ 1, p_i\ )$$

For each pathologist (PID), their career level (LEVEL), the specific diagnostic question (CATEGORY), the specimen type (SPECIMEN), and the subspecialty of the case (SPECIALTY), we used the logit link function and modeled the varying intercepts as follows:

The prior distribution for the intercepts and SD values were as follows:

$$\alpha_j \sim Normal\ (\ \square\ ,\ \sigma_\alpha\ ),\ \text{for}\ j = 1..17$$

$\beta_j \sim Normal\ (0, \sigma_\beta)$, for $j = 1..4$

$\gamma_j \sim Normal\ (0, \sigma_\gamma)$, for $j = 1..5$

$\delta_j \sim Normal\ (0, \sigma_\delta)$, for $j = 1..2$

$\varepsilon_j \sim Normal\ (0, \sigma_\varepsilon)$, for $j = 1..3$

$\sigma_\beta \sim Exponential\ (1)$

$\sigma_\gamma \sim Exponential\ (1)$

$\sigma_\delta \sim Exponential\ (1)$

$\sigma_\varepsilon \sim Exponential\ (1)$

The hyperpriors for the hyperparameters average pathologist and σ_α were set as follows:

$\sim Normal\ (0, 1.5)$

$\sigma_\alpha \sim Exponential\ (1)$

The SD value for was set at 1.5 since it produces a flat (weakly regularizing) prior after logit transformation [16,18]; moreover, we used an exponential distribution to model SD, because this assumes the least, for maximum entropy reasons [16,25-28], given the fact that σ is a nonnegative continuous parameter. To assess the validity of priors, we run prior predictive simulation of the model [16,29,30] (see Table S1 in Multimedia Appendix 1, and Multimedia Appendices 2 and 3). To limit divergent transitions, we reparametrized the models with a noncentered equivalent form [31,32]. Models were fit using Stan (a probabilistic programming language) and R [33,34]. Full anonymized data and custom code can be found in the public repository SmartCovid hosted on Github [35].

Study Survey

The survey was inspired by previous published works [36-38]. Briefly, the survey included 17 questions in a randomized order for all the pathologists, covering 3 fields: (1) attitude towards DP, (2) confidence in using DP solutions, and (3) satisfaction with DP. The survey was sent at the end of the digital experience. Pathologists were requested to answer the questions using a Likert scale, with scores ranging from 1 (strongly disagree) to 5 (strongly agree). The results were reported as the proportion of pathologists who assigned each single value of the Likert scale.

Results

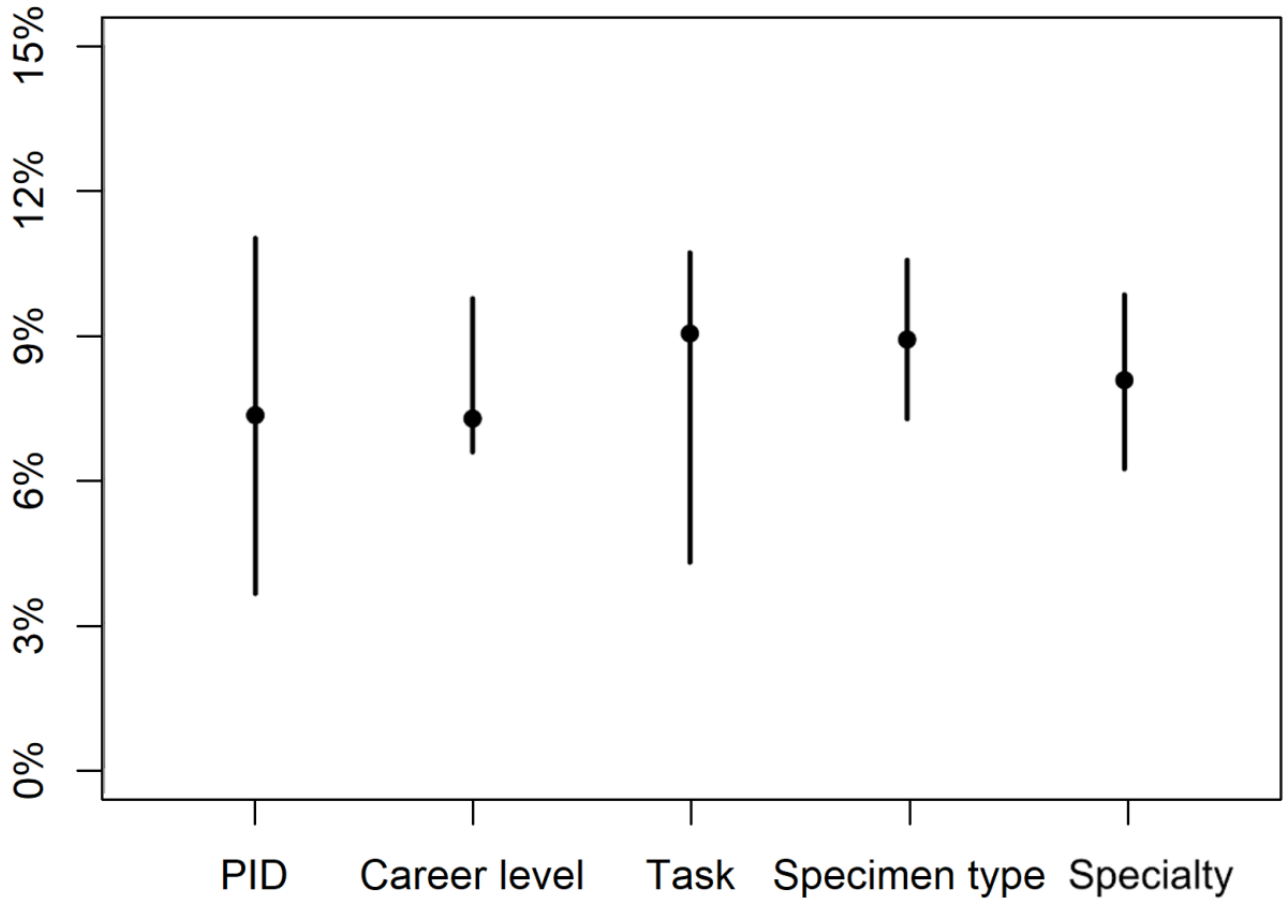
Quantitative Description

The pathologists answered 1345 of the total 1445 questions (100 missing answers in total), of which 1228 (91.30%) corresponded to the original diagnoses and were considered correct. Table 1 depicts the errors among each group of the 5 different categories recorded, and Figure 1 highlights the median (IQR) values of those categories. Considerable variation was observed among the performances of each pathologist, ranging from an error rate of 0.01 (1/67, Pathologist #4) to 0.32 (26/81, Pathologist #13), with a collective median error of 0.07 (IQR 0.04-0.11). This performance variation was tapered once the same data were considered after filtering among the different career levels, yielding the same median of 0.07, but a narrower IQR of 0.07-0.10. Moreover, some diagnostic tasks were more error prone than others; for instance, histotyping of the lesions had a very low rate of errors 0.01 (2/160), whereas grading was a more error-prone task with an error rate of 0.18 (27/147). The specimen type also resulted in different error rates, with surgical specimens easier to diagnose, with an error rate of 0.06 (40/716), than biopsy specimens, with a 2-fold error rate at 0.12 (77/629).

Table 1. Proportion of errors among different groups.

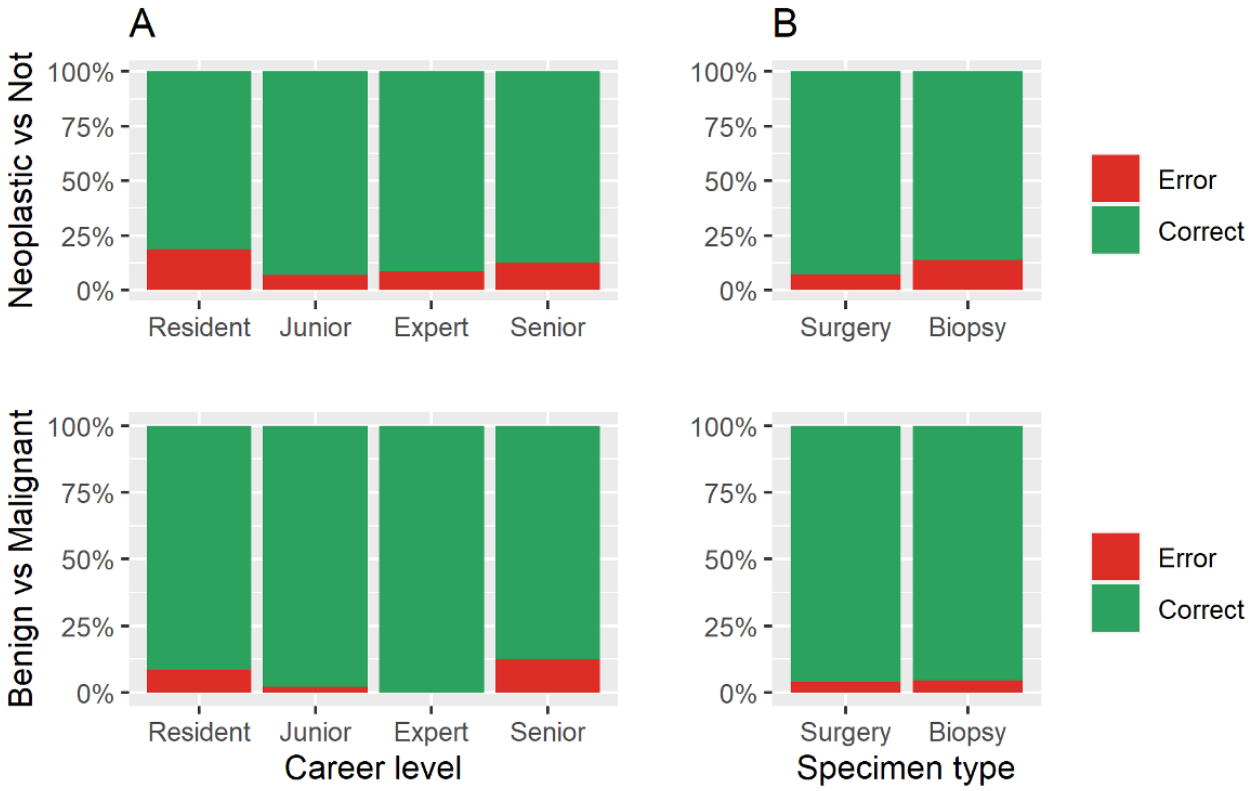
Group	Number of tasks performed	Number of errors	Error rate
Pathologist ID			
P1	84	5	0.06
P2	78	4	0.05
P3	82	7	0.09
P4	67	1	0.01
P5	82	7	0.09
P6	82	6	0.07
P7	83	2	0.02
P8	84	3	0.04
P9	82	5	0.06
P10	83	3	0.04
P11	82	9	0.11
P12	83	3	0.04
P13	81	26	0.32
P14	64	9	0.14
P15	84	12	0.14
P16	79	9	0.11
P17	65	6	0.09
Career level			
Resident	310	47	0.15
Junior	460	30	0.07
Expert	411	27	0.07
Senior	164	13	0.08
Category of the diagnostic task			
Neoplasia?	392	42	0.11
Malignant/benign?	258	11	0.04
Histopathological diagnosis?	388	35	0.09
Histotype?	160	2	0.01
Grade?	147	27	0.18
Specimen type			
Surgery	716	40	0.06
Biopsy	629	77	0.12
Case subspecialty			
Breast	550	64	0.12
Gastrointestinal	497	40	0.08
Urology	298	13	0.04
Total	1345	117	0.09

Figure 1. Error rates among different categories. This dot-bar plot depicts the median (IQR) values of error rates among different categories. The error rates showed the widest IQR among individual pathologists (PID), whereas the least IQR was noted for the career level and the specimen type (biopsy vs surgical).



Differences in error rates for two important tasks—differentiation between neoplastic and nonneoplastic processes and that between benign and malignant neoplastic processes—were observed among pathologists at different career levels and for different specimen types. The same error profile was observed across career levels, although the former task had a higher error rate (Figure 2A). However, even though the differentiation of a neoplastic process from a nonneoplastic one might be more challenging on a biopsy specimen, the distinction between a benign and malignant neoplasm was done with the same error rate regardless of the specimen type (Figure 2B). Differences in the prevalence of errors among individual pathologists and those at different career levels, as well as across diagnostic tasks, specimen type, and case subspecialty, are further highlighted in Multimedia Appendices 4 and 5.

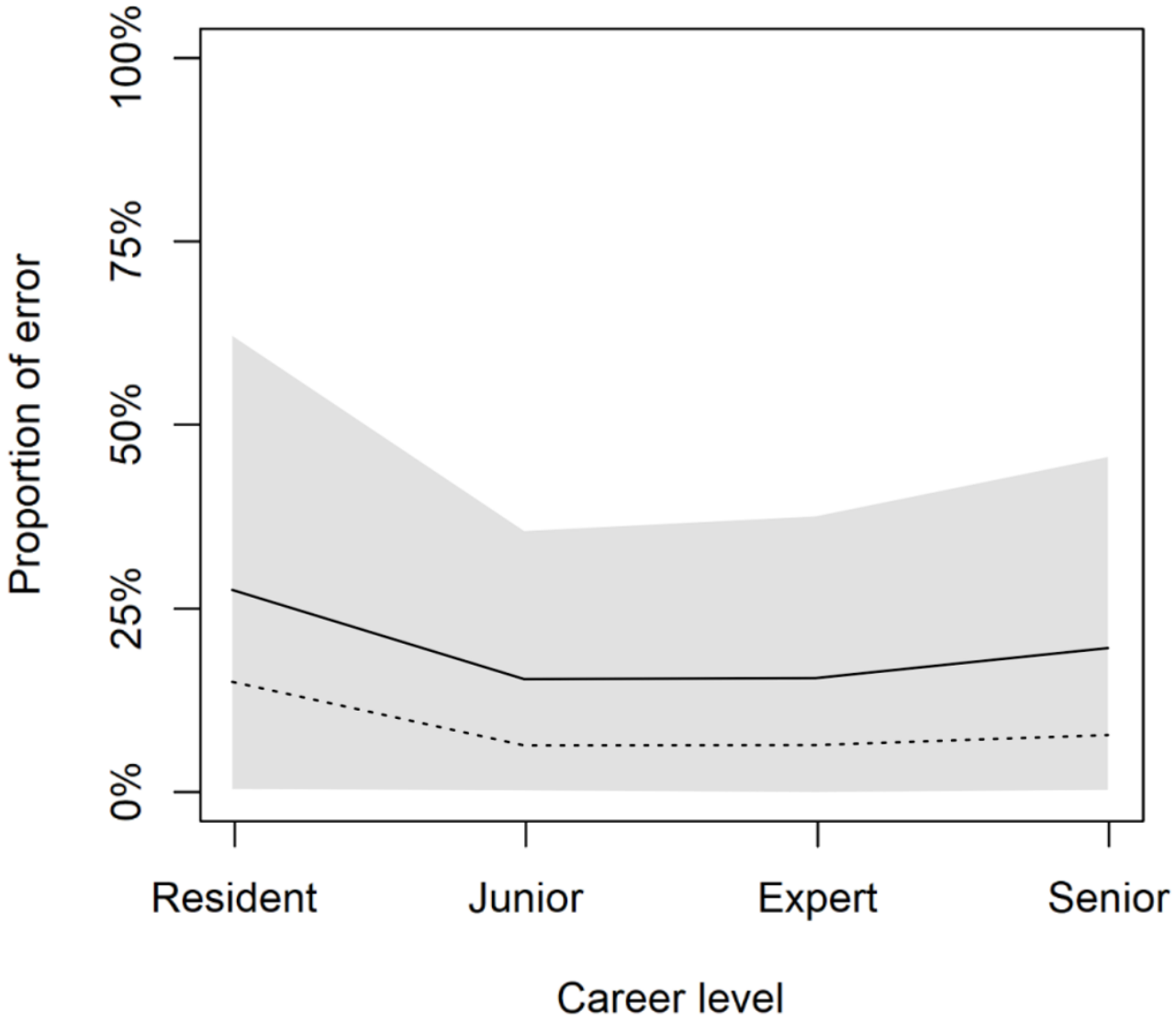
Figure 2. Raw proportion of errors across (A) career levels and (B) specimen types in performing two important tasks: differentiation between neoplastic and nonneoplastic processes and between malignant and benign tumors.



Prediction of Average Pathologist Performance

Diagnostics of the model's fit are shown in Multimedia Appendices 6, 7, and 8. The analysis reported a good overall performance: the average pathologist showed a negative mean coefficient of -1.8 with most of the posterior probability mass below 0 (given the model structure, positive values reflect the probability of making errors; Table S2 in Multimedia Appendix 1). The pathologists' individual performances and their career levels were the variables that showed less variance in predicting the error rate, whereas the specimen type, case subspecialty, and the particular type of task collectively showed more variance (Multimedia Appendix 9). Hence, we simulated the performance of an average pathologist at different career levels; this prediction shows better performance among pathologists at intermediate career levels of career (Figure 3).

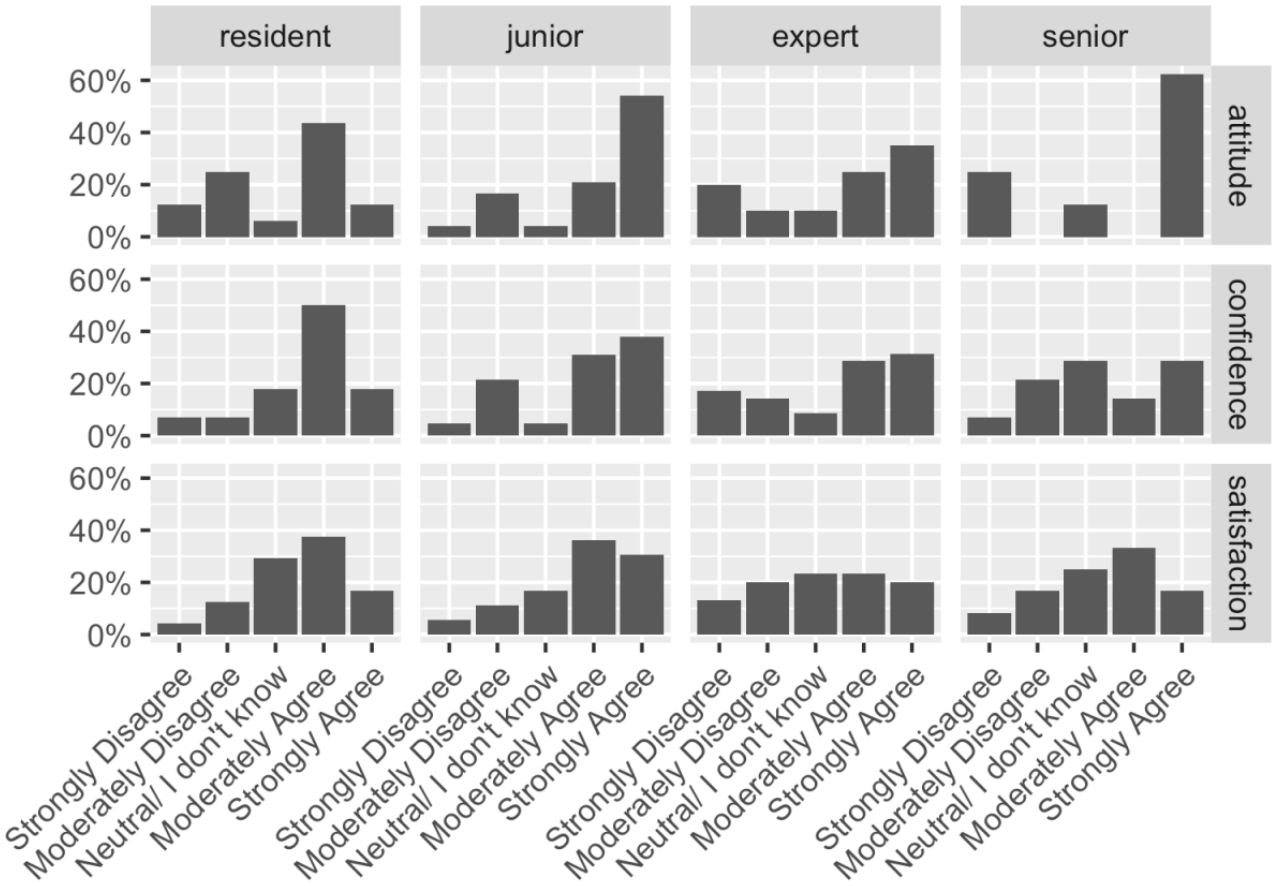
Figure 3. Prediction of average pathologist performance. Pathologists of intermediate levels of career perform better on average. The graph depicts the posterior predictive distributions for the multilevel model. Solid lines represent posterior mean values; shaded regions represent 89% high-posterior density interval; and dashed lines represent raw data.



Survey Results

Most pathologists reported a very good score (ie, 4 or 5 indicating they "moderately agree" and "strongly agree," respectively) for their attitude toward DP (44/68, 64%), confidence in DP (75/119, 63%), and satisfaction with DP (56/102, 54.9%). A detailed analysis of these parameters showed that the residents reported the highest value for confidence, junior pathologists reported the highest values for attitude and satisfaction, whereas expert and senior pathologists reported relatively lower levels of confidence in and satisfaction with DP (Figure 4).

Figure 4. Overview of the psychological aspect of the study. This series of graphs summarize the results of the survey conducted among pathologists at different career levels (residents, junior, expert, and senior) to evaluate their attitudes toward, confidence in, and satisfaction with digital pathology solutions.



Discussion

Principal Findings

Our study showed an overall discordant rate of 9% among diagnoses performed using digital slides and those performed using the gold standard (ie, conventional microscopy). However, when we considered the different diagnostic tasks, this rate dropped to less than 5% in the category "benign versus malignant tumor", which is probably the most clinically impacting category among the other diagnostic tasks. A systematic review of 38 pertinent studies published before 2015 reported a 7.6% overall discordance rate between digital and glass slide diagnoses. Among these studies, 17 studies reported a discordant rate higher than 5%, and 8 reported a discordant rate higher than 15% [39]. A later reanalysis of the same series fixed the overall discordance rate to 4% and major discrepancies to 1% [40]. A more recent review, covering studies published until 2018, reported a disagreement ranging from 1.7% to 13% [10]. Two multicentric, randomized, non-inferiority studies reported major discordant rates of 4.9% [41] and 3.6% [42] between diagnoses done by digital and glass slides. Furthermore, a study from a single, large academic center reported an overall diagnostic equivalency of 99.3% [43]. The same group was also the first to report about the use of DP during COVID-19 with an overall concordance of 98.8% [44]. Thus, despite our challenging approach to DP, the diagnostic performance we recorded was consistent with previous reports—a result that further supports the transition to DP.

In our study, a high proportion of errors was generated in small biopsy specimen type (12.2%) and diagnostic tasks involving tumor grading (23%). These results are consistent with those of the review by Williams et al [40]. The latter showed that 21% of all errors concerned grading or histotyping of malignant lesions, whereas 10% of the errors could be ascribed to the inability to find the target.

Moreover, recent studies have consistently reported high, intermediate, and low discordant rates for bladder, breast, and gastrointestinal tract specimens, respectively [41,42]—a finding suggesting intrinsic difficulties of specific areas. In contrast, we observed 4%, 8%, and 12% of discrepancies for urology, gastrointestinal tract, and breast specimens. This result could be attributed to a nonrandom selection of the cases and might represent a study limitation, biasing the value of the coefficients of specific parameters of the case subspecialty, similar to those of diagnostic tasks and the specimen type. However, these characteristics were excluded in the posterior predictive simulation, which was intended to represent how the different career levels might impact the pathologists' performance, after adjusting for all other factors.

As compared by the study by Hanna et al [44], our readiness to undertake digital diagnostic tasks was far from being mature in March 2020, and this study was specifically designed to identify

and illustrate the effects of such a sudden adoption of DP—something that had never been investigated before. Our results suggest that this abrupt transition might not influence the adoption of and performance with DP. However, different factors seem to be involved. In particular, data concerning major discrepancies between diagnoses using DP and gold standard methods disclosed an interesting feature. Both in the distinction of neoplastic versus non-neoplastic lesions and of benign versus malignant tumors, the worst results obtained were among residents and senior pathologists–2 contrasting categories in terms of pathologists' working experience. Therefore, these survey results might suggest an explanation to this paradoxical result: senior pathologists felt ready to diagnose a pathological process using a digital approach (ie, positive attitude) but were less prepared to use digital devices (ie, low confidence). Residents, in turn, had a high predisposition to using a digital device (ie, high confidence) but also had some concerns about diagnosis of a pathological process (ie, poor attitude). The hypothesis that senior pathologists were limited by a digital gap was supported by another finding: once they decided a lesion was malignant, they demonstrated the best performance with regard to tumor grading. By contrast, residents made several errors, likely due to their limited working experience. Lastly, even if expert pathologists showed a good diagnostic performance, they had the lowest level of satisfaction in DP. This result suggests that DP can be adopted rapidly for practical purposes. However, it also highlights a critical point of the process that needs to be addressed, possibly with adequate training or user-friendly equipment, and warrants further investigations.

Conclusions

Our study describes how the abrupt transition to DP affected the quality of diagnoses and qualitatively gauged the psychological aspects of this abrupt transition. Moreover, our study model highlighted the potential causes for these challenges and might inform what could be expected in other laboratories. In conclusion, the exceptional conditions dictated by the COVID-19 pandemic highlighted that DP could be adopted safely for diagnostic purposes by any skilled pathologist, even abruptly.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials and methods.
[DOCX File , 47 KB - jmir_v23i2e24266_app1.docx]

Multimedia Appendix 2
Coefficients of model parameters from the prior predictive simulation.
[PNG File , 116 KB - jmir_v23i2e24266_app2.png]

Multimedia Appendix 3
Simulation from the prior. This figure shows the meaning of the priors (ie, what the model thinks before it sees the data).
[PNG File , 88 KB - jmir_v23i2e24266_app3.png]

Multimedia Appendix 4
Proportion of errors among individual pathologists. Upper left panel shows the overall error rates. Upper right panel shows the error rates among different diagnostic tasks. Lower left panel shows the error rate among different specimen types. Lower right panel highlights the different error rates among different case subspecialties. GI: gastrointestinal, Uro: urology.
[PNG File , 143 KB - jmir_v23i2e24266_app4.png]

Multimedia Appendix 5
Proportion of errors among different career levels. Upper left panel shows the overall error rates. Upper right panel shows the error rates among the different diagnostic tasks. Lower left panel shows the error rate among different specimen types. Lower right panel highlights the different error rates among different case subspecialties. GI: gastrointestinal, Uro: urology.
[PNG File , 128 KB - jmir_v23i2e24266_app5.png]

Multimedia Appendix 6
Traceplot of the model fit - part A.
[PNG File , 276 KB - jmir_v23i2e24266_app6.png]

Multimedia Appendix 7
Traceplot of the model fit - part B.
[PNG File , 267 KB - jmir_v23i2e24266_app7.png]

Multimedia Appendix 8
Traceplot of the model fit - part C.
[PNG File , 111 KB - jmir_v23i2e24266_app8.png]

Multimedia Appendix 9
Model coefficients. Graphical representation of the coefficients for the model parameters conditional on the data. The lowest box depicts the coefficients for the hyper-parameter $\bar{\alpha}$ (alpha_bar) and the variances – the σ (sigma_a, b, [...] e) – of the categories of clusters modeled. All other boxes depict the distributions of the mean value for each element of the category considered. From top to bottom: the first box depicts the parameters of the pathologists' performance; the second, the parameters regarding the career level; the third, the diagnostic category analyzed; the fourth, the specimen type; and the fifth, the case subspecialty. Interpretation of the model at the parameter level is not possible because they combine in a very complicated way: prediction (ie, see how the model behave on the outcome scale, Figure 4 in the manuscript) is the only practical way to understand what the model "thinks".
[PNG File , 116 KB - jmir_v23i2e24266_app9.png]

References

1. Pantanowitz L, Sharma A, Carter A, Kurc T, Sussman A, Saltz J. Twenty years of digital pathology: an overview of the road travelled, what is on the horizon, and the emergence of vendor-neutral archives. J Pathol Inform 2018;9:40 [FREE Full text] [doi: 10.4103/jpi.jpi_69_18] [Medline: 30607307]
2. Griffin J, Treanor D. Digital pathology in clinical use: where are we now and what is holding us back? Histopathology 2017 Jan;70(1):134-145. [doi: 10.1111/his.12993] [Medline: 27960232]
3. Zarella MD, Bowman D, Aeffner F, Farahani N, Xthona A, Absar SF, et al. A practical guide to whole slide imaging: a white paper from the digital pathology association. Arch Pathol Lab Med 2019 Feb;143(2):222-234. [doi: 10.5858/arpa.2018-0343-ra]
4. Hartman D, Pantanowitz L, McHugh J, Piccoli A, OLeary M, Lauro G. Enterprise implementation of digital pathology: feasibility, challenges, and opportunities. J Digit Imaging 2017 Oct;30(5):555-560 [FREE Full text] [doi: 10.1007/s10278-017-9946-9] [Medline: 28116576]
5. Williams BJ, Treanor D. Practical guide to training and validation for primary diagnosis with digital pathology. J Clin Pathol 2020 Jul;73(7):418-422. [doi: 10.1136/jclinpath-2019-206319] [Medline: 31784420]
6. Stathonikos N, Nguyen TQ, Spoto CP, Verdaasdonk MAM, van Diest PJ. Being fully digital: perspective of a Dutch academic pathology laboratory. Histopathology 2019 Nov;75(5):621-635 [FREE Full text] [doi: 10.1111/his.13953] [Medline: 31301690]
7. Fraggetta F, Garozzo S, Zannoni G, Pantanowitz L, Rossi E. Routine digital pathology workflow: the Catania experience. J Pathol Inform 2017;8:51 [FREE Full text] [doi: 10.4103/jpi.jpi_58_17] [Medline: 29416914]
8. Retamero JA, Aneiros-Fernandez J, del Moral RG. Complete digital pathology for routine histopathology diagnosis in a multicenter hospital network. Arch Pathol Lab Med 2020 Feb;144(2):221-228. [doi: 10.5858/arpa.2018-0541-oa]
9. Thorstenson S, Molin J, Lundström C. Implementation of large-scale routine diagnostics using whole slide imaging in Sweden: Digital pathology experiences 2006-2013. J Pathol Inform 2014;5(1):14 [FREE Full text] [doi: 10.4103/2153-3539.129452] [Medline: 24843825]
10. Araújo ALD, Arboleda LPA, Palmier NR, Fonsêca JM, de Pauli Paglioni M, Gomes-Silva W, et al. The performance of digital microscopy for primary diagnosis in human pathology: a systematic review. Virchows Arch 2019 Mar;474(3):269-287. [doi: 10.1007/s00428-018-02519-z] [Medline: 30685784]
11. Gelman A, Carlin J. Beyond power calculations: assessing type S (sign) and type M (magnitude) errors. Perspect Psychol Sci 2014 Nov;9(6):641-651. [doi: 10.1177/1745691614551642] [Medline: 26186114]
12. Gelman A. The failure of null hypothesis significance testing when studying incremental changes, and what to do about it. Pers Soc Psychol Bull 2018 Jan;44(1):16-23. [doi: 10.1177/0146167217729162] [Medline: 28914154]
13. Gelman A. The problems with p-values are not just with p-values. Am Stat (Supplemental material to the ASA statement on p-values and statistical significance) 2016 Jun:129-133 [FREE Full text]
14. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models (Analytical Methods for Social Research). Cambridge: Cambridge University Press; 2006.
15. Gelman J, Carlin B, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. New York: CRC Press; 2013.
16. McElreath R. Statistical Rethinking: A Bayesian course with examples in R and Stan. Boca Raton: CRC Press; 2020.
17. Gelman A, Weakliem D. Of Beauty, Sex and Power - Too little attention has been paid to the statistical challenges in estimating small effects. American Scientist 2009;97(4):310. [doi: 10.1511/2009.79.310]
18. Renne SL, Valeri M, Tosoni A, Bertolotti A, Rossi R, Renne G, et al. Myoid gonadal tumor. Case series, systematic review, and Bayesian analysis. Virchows Arch 2020 Nov. [doi: 10.1007/s00428-020-02957-8] [Medline: 33140129]
19. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953 Jun;21(6):1087-1092. [doi: 10.1063/1.1699114]

20. Hoffman MD, Gelman A. The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res 2014 Apr;15:1593-1623 [FREE Full text]
21. Gelman A. Analysis of variance: why it is more important than ever. Ann Stat 2005;33(1):1-31 [FREE Full text]
22. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J Mach Learn Res 2010;11:3571-3594 [FREE Full text]
23. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput 2016 Aug 30;27(5):1413-1432. [doi: 10.1007/s11222-016-9696-4]
24. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. Stat Comput 2013 Aug 20;24(6):997-1016. [doi: 10.1007/s11222-013-9416-2]
25. Williams PM. Bayesian conditionalisation and the principle of minimum information. Br J Philos Sci 1980 Jun 01;31(2):131-144. [doi: 10.1093/bjps/31.2.131]
26. Caticha A, Giffin A. Updating probabilities. AIP Conference Proceedings 2006;872(1):31-42. [doi: 10.1063/1.2423258]
27. Giffin A. Maximum entropy: the universal method for inference. arXiv Preprint posted online January 20, 2009. [FREE Full text]
28. Jaynes T. The relation of Bayesian and maximum entropy method. In: Erickson GJ, Smith CR, editors. Maximum-Entropy and Bayesian Methods in Science and Engineering. Fundamental Theories of Physics (An International Book Series on The Fundamental Theories of Physics: Their Clarification, Development and Application). Dordrecht: Springer; 1988:29.
29. Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. Visualization in Bayesian workflow. J R Stat Soc Ser A 2019 Jan 15;182(2):389-402. [doi: 10.1111/rssa.12378]
30. Gelman A, Vehtari A, Simpson D, Margossian CC, Carpenter B, Yao Y, et al. Bayesian Workflow. arXiv Preprint posted online November 3, 2020. [FREE Full text]
31. Papaspiliopoulos O, Roberts GO, Sköld M. A general framework for the parametrization of hierarchical models. Statist Sci 2007 Feb;22(1):59-73. [doi: 10.1214/088342307000000014]
32. 22.7 Reparameterization. Stan Development Team Stan User's Guide Version 2. URL: https://mc-stan.org/docs/2_25/stan-users-guide/reparameterization-section.html [accessed 2020-12-09]
33. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: a probabilistic programming language. J Stat Soft 2017;76(1):1-32. [doi: 10.18637/jss.v076.i01]
34. The R Project for Statistical Computing. The R Foundation. 2019. URL: https://www.r-project.org/ [accessed 2021-01-28]
35. SmartCovid: Datasets and code for the study. GitHub. URL: https://github.com/slrenne/SmartCovid [accessed 2021-01-29]
36. Randell R, Ruddle RA, Treanor D. Barriers and facilitators to the introduction of digital pathology for diagnostic work. Stud Health Technol Inform 2015;216:443-447. [Medline: 26262089]
37. Pavone F. Guida rapida per operatori in campo contro il COVID-19: Autovalutazione dello stress e Gestione del disagio emotivo. 2020 Mar 29. URL: https://associazioneitalianacasemanager.it/wp-content/uploads/2020/04/COVID_19_e_stress_professionale_3_1-FP-ASST.pdf [accessed 2020-12-09]
38. Retamero JA, Aneiros-Fernandez J, Del Moral RG. Microscope? No, thanks: user experience with complete digital pathology for routine diagnosis. Arch Pathol Lab Med 2020 Jun;144(6):672-673 [FREE Full text] [doi: 10.5858/arpa.2019-0355-LE] [Medline: 32459536]
39. Goacher E, Randell R, Williams B, Treanor D. The diagnostic concordance of whole slide imaging and light microscopy: a systematic review. Arch Pathol Lab Med 2017 Jan;141(1):151-161 [FREE Full text] [doi: 10.5858/arpa.2016-0025-RA] [Medline: 27399211]
40. Williams BJ, DaCosta P, Goacher E, Treanor D. A systematic analysis of discordant diagnoses in digital pathology compared with light microscopy. Arch Pathol Lab Med 2017 Dec;141(12):1712-1718 [FREE Full text] [doi: 10.5858/arpa.2016-0494-OA] [Medline: 28467215]
41. Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG, et al. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology. Am J Surg Pathol 2017:1. [doi: 10.1097/pas.0000000000000948]
42. Borowsky AD, Glassy EF, Wallace WD, Kallichanda NS, Behling CA, Miller DV, et al. Digital whole slide imaging compared with light microscopy for primary diagnosis in surgical pathology. Arch Pathol Lab Med 2020 Oct 01;144(10):1245-1253 [FREE Full text] [doi: 10.5858/arpa.2019-0569-OA] [Medline: 32057275]
43. Hanna MG, Reuter VE, Hameed MR, Tan LK, Chiang S, Sigel C, et al. Whole slide imaging equivalency and efficiency study: experience at a large academic center. Mod Pathol 2019 Jul;32(7):916-928. [doi: 10.1038/s41379-019-0205-0] [Medline: 30778169]
44. Hanna MG, Reuter VE, Ardon O, Kim D, Sirintrapun SJ, Schüffler PJ, et al. Validation of a digital pathology system including remote review during the COVID-19 pandemic. Mod Pathol 2020 Nov;33(11):2115-2127 [FREE Full text] [doi: 10.1038/s41379-020-0601-5] [Medline: 32572154]

Abbreviations

DP: digital pathology

Edited by G Eysenbach; submitted 11.09.20; peer-reviewed by R Poluru, B Kaas-Hansen; comments to author 01.12.20; revised version received 09.12.20; accepted 14.12.20; published 22.02.21.

Please cite as:
Giaretto S, Renne SL, Rahal D, Bossi P, Colombo P, Spaggiari P, Manara S, Sollai M, Fiamengo B, Brambilla T, Fernandes B, Rao S, Elamin A, Valeri M, De Carlo C, Belsito V, Lancellotti C, Cieri M, Cagini A, Terracciano L, Roncalli M, Di Tommaso L
Digital Pathology During the COVID-19 Outbreak in Italy: Survey Study
J Med Internet Res 2021;23(2):e24266
URL: https://www.jmir.org/2021/2/e24266
doi:10.2196/24266
PMID:33503002

Original Paper

Learning From Past Respiratory Infections to Predict COVID-19 Outcomes: Retrospective Study

Shengtian Sang[1*], PhD; Ran Sun[1*], PhD, RN; Jean Coquet[1*], PhD; Harris Carmichael[2], MD; Tina Seto[3], MSc; Tina Hernandez-Boussard[1], PhD, MPH, MSc

[1]Department of Medicine, Biomedical Informatics, Stanford University, Stanford, CA, United States

[2]Intermountain Health, Salt Lake City, UT, United States

[3]Technology and Digital Solutions, Stanford University, Stanford, CA, United States

[*]these authors contributed equally

Corresponding Author:
Tina Hernandez-Boussard, PhD, MPH, MSc
Department of Medicine
Biomedical Informatics
Stanford University
1265 Welch Rd, 245
Stanford, CA, 94305-5479
United States
Phone: 1 650 725 5507
Email: boussard@stanford.edu

Abstract

Background: For the clinical care of patients with well-established diseases, randomized trials, literature, and research are supplemented with clinical judgment to understand disease prognosis and inform treatment choices. In the void created by a lack of clinical experience with COVID-19, artificial intelligence (AI) may be an important tool to bolster clinical judgment and decision making. However, a lack of clinical data restricts the design and development of such AI tools, particularly in preparation for an impending crisis or pandemic.

Objective: This study aimed to develop and test the feasibility of a "patients-like-me" framework to predict the deterioration of patients with COVID-19 using a retrospective cohort of patients with similar respiratory diseases.

Methods: Our framework used COVID-19–like cohorts to design and train AI models that were then validated on the COVID-19 population. The COVID-19–like cohorts included patients diagnosed with bacterial pneumonia, viral pneumonia, unspecified pneumonia, influenza, and acute respiratory distress syndrome (ARDS) at an academic medical center from 2008 to 2019. In total, 15 training cohorts were created using different combinations of the COVID-19–like cohorts with the ARDS cohort for exploratory purposes. In this study, two machine learning models were developed: one to predict invasive mechanical ventilation (IMV) within 48 hours for each hospitalized day, and one to predict all-cause mortality at the time of admission. Model performance was assessed using the area under the receiver operating characteristic curve (AUROC), sensitivity, specificity, positive predictive value, and negative predictive value. We established model interpretability by calculating SHapley Additive exPlanations (SHAP) scores to identify important features.

Results: Compared to the COVID-19–like cohorts (n=16,509), the patients hospitalized with COVID-19 (n=159) were significantly younger, with a higher proportion of patients of Hispanic ethnicity, a lower proportion of patients with smoking history, and fewer patients with comorbidities (P<.001). Patients with COVID-19 had a lower IMV rate (15.1 versus 23.2, P=.02) and shorter time to IMV (2.9 versus 4.1 days, P<.001) compared to the COVID-19–like patients. In the COVID-19–like training data, the top models achieved excellent performance (AUROC>0.90). Validating in the COVID-19 cohort, the top-performing model for predicting IMV was the XGBoost model (AUROC=0.826) trained on the viral pneumonia cohort. Similarly, the XGBoost model trained on all 4 COVID-19–like cohorts without ARDS achieved the best performance (AUROC=0.928) in predicting mortality. Important predictors included demographic information (age), vital signs (oxygen saturation), and laboratory values (white blood cell count, cardiac troponin, albumin, etc). Our models had class imbalance, which resulted in high negative predictive values and low positive predictive values.

Conclusions: We provided a feasible framework for modeling patient deterioration using existing data and AI technology to address data limitations during the onset of a novel, rapidly changing pandemic.

(J Med Internet Res 2021;23(2):e23026) doi:10.2196/23026

KEYWORDS

COVID-19; invasive mechanical ventilation; all-cause mortality; machine learning; artificial intelligence; respiratory; infection; outcome; data; feasibility; framework

Introduction

The SARS-CoV-2 virus, which causes the disease COVID-19, has infected almost 107 million people worldwide and is responsible for more than 2.3 million deaths [1]. Around the globe, patients with COVID-19 have a broad range of symptoms and disease severities. However, patients with COVID-19 demonstrate several symptoms that are striking in their commonality and nonspecificity to other well-known respiratory infections, such as pneumonia, influenza, and acute respiratory distress syndrome (ARDS) [2-5]. In fact, initial reporting of the disease indicated that most patients present with pneumonia- or influenza-like illnesses [6,7].

COVID-19 shares symptoms with other respiratory illnesses; from these diseases, we can learn about the clinical progression of patients, including the progression of patients presenting with severe hypoxemia followed by rapid deterioration, which often requires advanced life support such as invasive mechanical ventilation (IMV). Worldwide, the pandemic has left health systems struggling with capacity limits, especially regarding intensive care units (ICU) and mechanical ventilators [8,9]. Patients with COVID-19 remain on mechanical ventilation for an average of 10 to 20 days, further limiting the availability of this scarce resource [10,11]. Furthermore, an earlier study showed a high mortality rate of up to about 60% among critically ill patients [12]. The severity of cases has put great pressure on health systems, leading to a shortage of intensive care resources. Thus, understanding who will become critically ill and consume scarce resources as a result of this emerging disease can improve resource, hospital, and societal planning, and increase the number of lives saved.

However, given the novelty of COVID-19 and lack of clinical experience, the health care community is grappling for robust clinical data to learn about the disease. Under the appropriate assumptions, artificial intelligence (AI) may help with the planning of COVID-19 responses and guide clinical decisions in this time of uncertainty. Given the number of patients that may become infected with COVID-19, it is essential to understand who will be more likely to develop severe illness and need scarce resources at the time of presentation to the health care system, and AI can help with this determination. Indeed, AI technologies related to COVID-19 outcomes are emerging, as indicated in a recent published systematic review, with a majority of these studies conducted in China [13]. Predictive models using rule-based scoring tools and machine learning approaches have been applied to predict clinical deterioration in hospitalized patients and support health care providers in triaging patients when resources are limited [14-19]. However, there are many concerns and barriers to making this a reality [20], including uncertainty in the risk factors associated with disease progression, a limited number of patients whose data can be used to train and test models, and no public data sets available to test and validate models outside of a single health care setting. Of the COVID-19 AI studies published thus far, many have been designed and developed retrospectively on a small number of patient cases, limiting their validity and generalizability to other populations [15,21,22].

To help guide clinical decisions during the COVID-19 pandemic, we developed a framework to bootstrap AI models for outcomes of patients with COVID-19 using COVID-19–like cohorts to develop and train AI models to predict IMV within 48 hours and mortality using features associated with outcomes of patients with COVID-19. The COVID-19–like cohorts included patients diagnosed with bacterial pneumonia, influenza, viral pneumonia, and ARDS between 2008-2019. We tested the models' performances on hospitalized patients with COVID-19. This framework may be particularly important in a novel and accelerated outbreak where clinicians and health care systems are forced to make difficult decisions without past experience of the specific disease at hand. In the void created by a lack of clinical experience with COVID-19, AI trained with data from COVID-19–like disorders may be an important way to bolster clinical judgment and decision making.

Methods

Study Design and Data Source

This retrospective study used data from electronic health records (EHRs) of patients admitted to the Stanford Healthcare Alliance (SHA) from January 1, 2008, to July 11, 2019. SHA is an integrated health system that includes an academic hospital, a community hospital, and a primary/specialty health care alliance. This study received approval from the institute's Institutional Review Board (IRB). All source codes for this work are available at the Stanford Digital Repository [23].

Study Cohorts

This study included two retrospective cohorts: a COVID-19–like cohort (n=16,509) and a COVID-19 cohort (n=159).

COVID-19–like Cohort

The COVID-19–like cohort used for model training included patients hospitalized with a diagnosis of bacterial pneumonia, influenza, viral pneumonia, unspecified pneumonia, or ARDS, using International Classification of Diseases codes (ICD-9 and ICD-10; Table S1 in Multimedia Appendix 1), between January 1, 2008, and July 11, 2019. These diseases were selected because of their similarity to COVID-19 in clinical manifestation, histological features, and disease progression [4,7,24]. Patients

with complete missing lab data were excluded from the study (n=1712, 10.4%). For patients with multiple ICD codes for different conditions, the following rule was applied for disease categorization: influenza → viral pneumonia → bacterial pneumonia → unspecified pneumonia. Among these COVID-19–like cohorts, those who developed ARDS formed a separate COVID-19–like cohort.

COVID-19 Cohort

We included adults with a confirmed COVID-19 diagnosis who were hospitalized between March 1, 2020, and July 11, 2020. A confirmed COVID-19 diagnosis was defined as either a positive SARS-CoV-2 RNA detection test or a diagnosis code for COVID-19 (Table S1 in Multimedia Appendix 1). All patients were observed throughout their hospital encounter.

Data Collection

Patient demographics and clinical information were captured from EHRs. We selected the most relevant features identified from the literature [2,3,5], including demographics, existing comorbid conditions, smoking history, symptoms at initial presentation, coinfection with other respiratory pathogens, and laboratory values (Textboxes 1-3). The patient's existing comorbidities, including cardiovascular disease, diabetes, cancer, hypertension, chronic respiratory disease, respiratory failure, and kidney disease, were determined over a 3-year period prior to hospital admission for the retrospective cohort. Laboratory values 2 weeks prior to and during the hospital stay were extracted for the retrospective and COVID-19 cohorts. Laboratory values on the day of IMV were excluded to ensure the values were not taken after IMV.

Textbox 1. Features included in the invasive mechanical ventilation and mortality model development (demographics and clinical characteristics).

Demographics and clinical characteristics

- Age at admission (years)
- Gender
- Race/ethnicity
- Ever smoked (all life before admission)

Comorbidity present (3 years before admission)

- Cancer
- Chronic respiratory disease
- Cardiovascular disease
- Hypertension
- Type 2 diabetes
- Respiratory failure
- Kidney disease
- Alzheimer disease
- Cirrhosis

Symptoms (15 days before admission)

- Cough
- Dyspnea
- Tachypnea (respiratory rate >20)
- Hypoxemia (oxygen saturation ≤90%)
- Rhinorrhea
- Nose congestion
- Fever (temperature >37 C/98.6 F)
- Sputum
- Pharyngitis (sore throat)
- Headache
- Fatigue
- Conjunctivitis
- Diarrhea
- Anosmia
- Myalgias

Textbox 2. Features included in the invasive mechanical ventilation and mortality model development (laboratory findings).

Laboratory findings

- White blood cell count, K/μL
- Lymphocyte count, K/μL
- Alanine aminotransferase, U/L
- Aspartate aminotransferase, U/L
- Creatinine, mg/dL
- Lactate dehydrogenase, U/L
- Creatine kinase, U/L
- Cardiac troponin, ng/mL
- D-dimer, ng/mL
- Prothrombin time, seconds
- Serum ferritin, ng/mL
- Procalcitonin, ng/mL
- Platelet count, K/μL
- C-reactive protein, mg/dL
- Total bilirubin, mg/dL
- Blood urea nitrogen, mg/dL
- Albumin, g/dL
- Oxygen saturation, mm Hg
- Fraction of inspired oxygen, %
- Sodium, mmol/L
- Potassium, mmol/L

Textbox 3. Features included in the invasive mechanical ventilation and mortality model development (coinfections).

Coinfection (15 days before admission)

- Adenovirus
- Chlamydia pneumoniae
- Coronavirus
- Influenza A
- Influenza B
- Metapneumovirus
- Mycoplasma pneumonia
- Parainfluenza 1
- Parainfluenza 2
- Parainfluenza 3
- Parainfluenza 4
- Rhinovirus/enterovirus
- Respiratory syncytial virus

Outcome

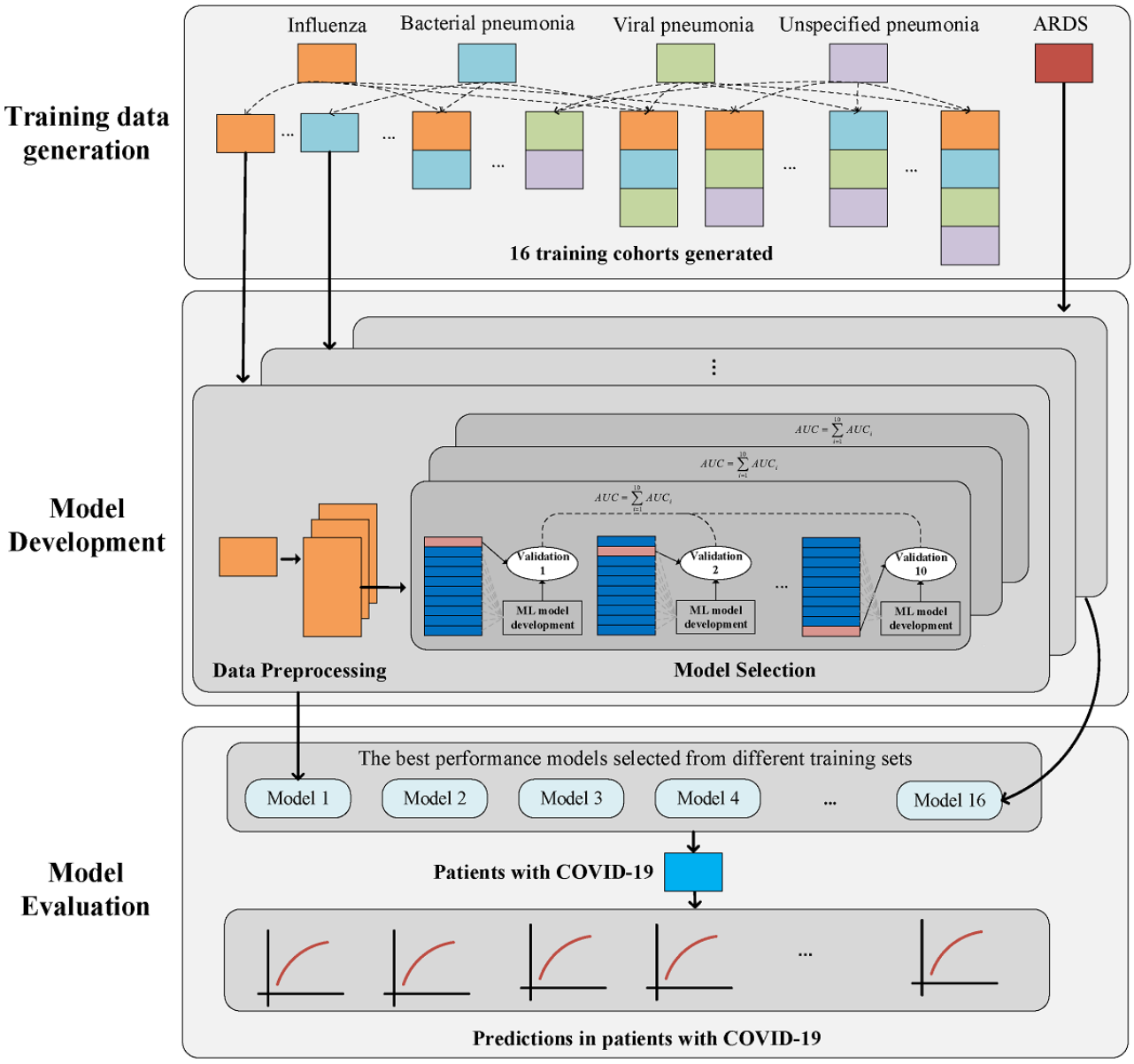
The outcome of interest was IMV in the next 48 hours and all-cause mortality. In the COVID-19–like cohort, the need for IMV in 48 hours was identified from the EHRs using the presence of IMV date and time. This time point was selected due to the rapid deterioration of disease conditions observed in patients with COVID-19. Deceased individuals who did not receive IMV during their hospital stay were considered nonintubated patients. All-cause mortality was identified using

the death date recorded in the EHRs, including those who died during their hospital stay and within two weeks of hospital discharge.

Artificial Intelligence Framework

We developed a three-step framework to bootstrap machine learning models to predict IMV and mortality among hospitalized patients with COVID-19. Figure 1 illustrates the study design framework, including training data generation, model development, and model evaluation.

Figure 1. The methodological framework for selecting training data cohort, model development, and model evaluation for predicting intubation and mortality.



Training Data Generation

To determine the most appropriate COVID-19–like cohort for the prediction task, we explored different combinations of the 4 COVID-19–like cohorts (influenza, bacterial pneumonia, viral pneumonia, and unspecified pneumonia), including any single disease, any two diseases, any three diseases, and all four diseases combined to construct 15 training cohorts. Due to the differences between patients with COVID-19 and patients with ARDS, these COVID-19–like patients who developed ARDS were considered as an additional exploratory cohort to examine whether or not it could contribute to better prediction for the patients with COVID-19. Since we selected the best-performing machine learning model from each COVID-19–like cohort, 16 models were developed from the 16 COVID-19–like cohorts.

Model Development

Data Preprocessing

Features with missing values in the training set were imputed as –1. Due to the unbalanced data in training cohorts (Table S2 in Multimedia Appendix 1), oversampling approaches were used to generate synthetic samples for the minority class and to balance the positive and negative training set. These approaches included synthetic minority oversampling technique

(SMOTE), borderline SMOTE, support vector machine (SVM) SMOTE, and random undersampling methods. Random undersampling was initially applied to trim the number of examples in the majority class to twice that of the minority group (positive cases), then the oversampling method was used to synthesize the minority class to balance the class distribution [25].

Model Training

Each of the 16 preprocessed COVID-19–like cohorts were split into 70% training and 30% validation sets with stratified random sampling. We derived four machine learning algorithms, including SVM and three tree-based ensemble algorithms (decision tree, AdaBoost, and XGBoost). We selected three decision tree–based algorithms because they have previously been applied to predict clinical events in patients with respiratory diseases based on EHR data [16,26,27]. We included models that were frequently applied for clinical prediction of severe patient outcomes [16,26,28]. In total, two steps were involved in model training: (1) using the training data set, a 10-fold cross-validation strategy was used to train the machine learning models, while grid search technique was used to search all combinations of hyperparameters and determine the best hyperparameters, and (2) using all training data, the models were retrained with the best hyperparameters (obtained in step 1). The validation data were used to monitor the performance of the model to avoid overfitting in training data. The final model was derived when the performance of the model on the validation data set did not improve after 20 training iterations. The detailed processes of model training are presented in Figure S1 in Multimedia Appendix 1.

Model Evaluation

COVID-19–like Cohorts

The performance of the models on the training cohorts was compared using the area under the receiver operating characteristic curve (AUROC). The default threshold of 0.5 was selected for interpreting probabilities to class labels. For IMV, to be a true positive, the model had to predict the need for IMV within 48 hours with a risk score of ≥0.5, and the patient had to receive IMV within this time interval. If the risk score was ≥0.5 and the patient did not receive IMV in 48 hours, the patient was treated as a false positive. The inverse strategy was applied for true negative and false negative cases. For mortality, patients who died during the hospital course or within two weeks after discharge with a risk score of ≥0.5 were considered as true positives. The inverse strategy was applied for the true negative and false negative mortality cases. To interpret the model, feature importance for predicting IMV and mortality were presented using SHapley Additive exPlanations (SHAP) values [29].

COVID-19 Cohort

The best training cohort was selected based on model performance (ie, AUROC) on the COVID-19 cohort. To determine the feasibility of models trained on COVID-19–like cohorts to identify patients at high risk of IMV within 48 hours and all-cause mortality, we calculated the AUROC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for patients with COVID-19.

Statistical Analysis

Descriptive statistics were used to compare the characteristics of the COVID-19–like and COVID-19 populations. An independent *t* test or Mann-Whitney *U* test were used wherever appropriate for comparing continuous features. The Pearson chi-square test was used for categorical features, and the Fisher exact test was used when the number in the cell was <5. Algorithms were developed using the training cohort and assessed on the independent validation cohort, which played no role in model development, by calculating the PPV and AUROC. Algorithms were further tested on the independent COVID-19 cohort. A threshold of 0.50 was set for each model, and PPV and all other threshold-dependent performance metrics were derived at this alert rate. As the PPV is threshold dependent, AUROC was also compared among models. We chose to present AUROC values because they are a threshold-independent measure of discrimination. Statistical significance for primary analysis was set at $P<.05$. All tests were two-tailed.

Results

Cohort Description

There were a total of 16,509 patients in the COVID-19–like cohorts, and 159 patients in the COVID-19 cohort (Table 1). Compared to the COVID-19–like cohorts with pneumonia and/or influenza, patients with COVID-19 were significantly younger (mean 56.9 [SD 20.3] years versus mean 65.8 [SD 19.3] years, $P<.001$), less often White (27% versus 58.2%, $P<.001$), and more often Hispanic (46.5% versus 13.6%, $P<.001$). In addition, patients with COVID-19 had a significantly smaller proportion of ever-smokers (29.6% versus 41.3%, $P=.003$) and fewer patients with COVID-19 had comorbidities (37.1% versus 67.1%, $P<.001$). There were significant differences in IMV rate and time to IMV between the COVID-19–like patients and patients with COVID-19; the patients with COVID-19 had a lower IMV rate (15.1% versus 23.2%, $P=.02$) and shorter time to IMV (2.9 [SD 3.9] days versus 4.1 [SD 7.3] days, $P<.001$). Additionally, we observed a significantly lower mortality rate in patients with COVID-19 compared to the COVID-19–like cohorts ($P<.001$).

Table 1. Comparison of patient demographics in COVID-19–like and COVID-19 cohorts[a].

Characteristics	COVID-19–like cohort (n=16,509)						COVID-19 cohort (n=159)	*P* value[b]
	Influenza	Bacterial pneumonia	Viral pneumonia	Unspecified pneumonia	Acute respiratory distress syndrome	Total		
Total, n (%)	1076 (6.5)	2779 (16.8)	681 (4.1)	10,308 (62.4)	1665 (10.1)	16,509	159	
Age (years), mean (SD)	63.9 (20.7)	63.9 (19.5)	63.8 (19.7)	67.2 (18.9)	62.3 (18.8)	65.8 (19.3)	56.9 (20.3)	<.001
Gender, n (%)								.08
Female	557 (51.8)	1212 (43.6)	346 (50.8)	4823 (46.8)	730 (43.8)	7668 (46.4)	85 (53.5)	
Male	519 (48.2)	1567 (56.4)	335 (49.2)	5485 (53.2)	933 (56.0)	8839 (53.5)	74 (46.5)	
Race, n (%)								<.001
White	585 (54.4)	1674 (60.2)	376 (55.2)	6096 (59.1)	872 (52.4)	9603 (58.2)	43 (27.0)	
Asian	178 (16.5)	429 (15.4)	123 (18.1)	1640 (15.9)	284 (17.1)	2654 (16.1)	23 (14.5)	
Black	72 (6.7)	130 (4.7)	38 (5.6)	598 (5.8)	110 (6.6)	948 (5.7)	<10 (<6.3)	
Other	241 (22.4)	546 (19.6)	144 (21.2)	1974 (19.2)	399 (24.0)	3304 (20.0)	76 (47.8)	
Ethnicity, n (%)								<.001
Non-Hispanic	898 (83.5)	2321 (83.5)	568 (83.4)	8786 (85.2)	1335 (80.2)	13,908 (84.2)	81 (50.9)	
Hispanic	170 (15.8)	380 (13.7)	106 (15.6)	1325 (12.9)	267 (16.0)	2248 (13.6)	74 (46.5)	
Insurance, n (%)								<.001
Private	263 (24.4)	605 (21.8)	176 (25.8)	2160 (21.0)	414 (24.9)	3618 (21.9)	32 (20.1)	
Public	747 (69.4)	1974 (71.0)	477 (70.0)	7447 (72.5)	1129 (67.8)	11,804 (71.5)	101 (63.5)	
Other	66 (6.2)	200 (7.2)	28 (4.1)	671 (6.5)	122 (7.4)	1087 (6.6)	26 (16.4)	
Ever smoked, n (%)	402 (37.4)	1128 (40.6)	283 (41.6)	4375 (42.4)	635 (38.1)	6823 (41.3)	47 (29.6)	.003
At least one comorbidity[c]	771 (71.7)	1745 (62.8)	571 (83.8)	7000 (67.9)	997 (59.9)	11,084 (67.1)	59 (37.1)	<.001
Invasive mechanical ventilation, n (%)	91 (8.5)	1031 (37.1)	95 (14.0)	1811 (17.6)	797 (47.9)	3825 (23.2)	24 (15.1)	.02
Time to invasive mechanical ventilation (days), mean (SD)	4.4 (7.2)	3.8 (6.7)	6.9 (12.9)	4.0 (7.2)	4.3 (7.4)	4.1 (7.3)	2.9 (3.9)	<.001
All-cause mortality, n (%)	62 (5.8)	513 (18.5)	82 (12.0)	1586 (15.4)	541 (32.5)	2789 (16.9)	<10 (<6.3)	<.001

[a]Percentages may not add up to 100 due to missing values.

[b]The *P* value was calculated between the total of 4 retrospective COVID-19–like cohorts and the COVID-19 cohort.

[c]Comorbidities include cardiovascular disease, diabetes, cancer, hypertension, chronic respiratory disease, respiratory failure, kidney disease, Alzheimer disease, and cirrhosis.

Tables 2 and 3 compare symptoms at presentation and laboratory results during a hospital course between the COVID-19–like cohort and patients with COVID-19. Patients with COVID-19 experience a wider range and greater number of symptoms compared to the COVID-19–like cohort with pneumonia and influenza ($P<.05$). The average clinical laboratory results during hospitalization were also different (Table 3). The median laboratory values of D-dimer, lactate dehydrogenase, and ferritin elevated in hospitalized patients with COVID-19. Compared to the retrospective cohort of hospitalized patients with pneumonia and influenza, patients with COVID-19 had significantly higher values for lymphocyte count, platelet count, aspartate aminotransferase, albumin, lactate dehydrogenase, and ferritin, and lower values for white blood cell count, bilirubin, blood urea nitrogen, D-dimer, procalcitonin, creatinine, and prothrombin time.

Table 2. Comparison of symptoms at admission of the COVID-19–like cohort and patients with COVID-19.

Symptoms at admission	COVID-19–like cohort (n=16,509), n (%)	COVID-19 cohort (n=159), n (%)	*P* value
Cough	2643 (16.0)	116 (73.0)	<.001
Dyspnea	6087 (36.9)	119 (74.8.)	<.001
Fever	3147 (19.1)	103 (64.8)	<.001
Fatigue	2046 (12.4)	67 (42.1)	<.001
Myalgias	243 (1.5)	52 (32.7)	<.001
Hypoxemia	3910 (23.7)	51 (32.1)	.01
Headache	638 (3.9)	28 (17.6)	<.001
Diarrhea	1784 (10.8)	29 (18.2)	.003
Tachypnea	821 (5.0)	57 (35.8)	<.001
Pharyngitis (sore throat)	53 (0.3)	28 (17.6)	<.001
Sputum	0 (0.0)	26 (16.4)	N/A
Rhinorrhea	0 (0.0)	6 (3.8)	N/A
Nasal congestion	62 (0.4)	10 (6.3)	<.001
Anosmia	8 (0.0)	20 (12.6)	N/A

Table 3. Comparison of laboratory values of the COVID-19–like cohort and patients with COVID-19.

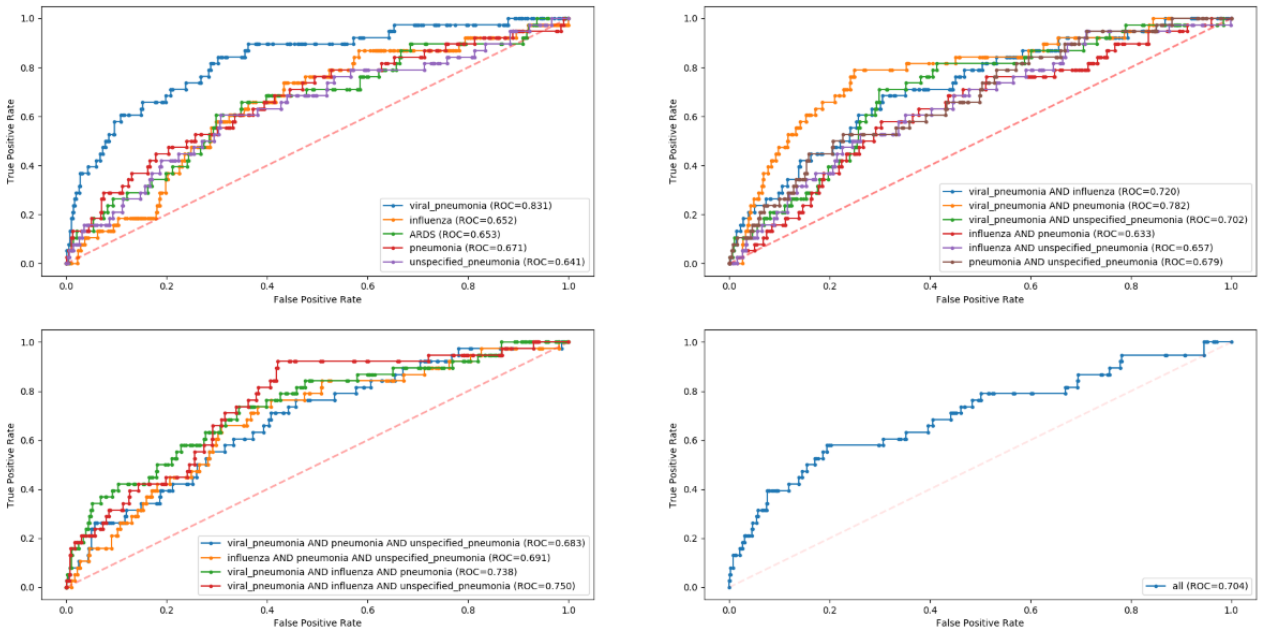
Laboratory values	Values missing, n (%)	Median (IQR)	Values missing, n (%)	Median (IQR)	*P* value
White blood cell count, K/μL	2767 (16.8)	8.9 (5.7)	0 (0.0)	6.8 (3.5)	<.001
Lymphocyte count, K/μL	3386 (20.5)	1.0 (0.8)	2 (1.3)	1.1 (0.7)	.008
Platelet count, K/μL	2768 (16.8)	197.3 (134.5)	0 (0.0)	240.9 (110.7)	<.001
Alanine aminotransferase, units/L	3977 (24.1)	32.0 (25.5)	1 (0.6)	33.2 (39.7)	.65
Aspartate aminotransferase, units/L	3803 (23.0)	29.0 (24.9)	1 (0.6)	40.9 (28.6)	<.001
Total bilirubin, mg/dL	5125 (31.0)	0.6 (0.5)	2 (1.3)	0.4 (0.3)	<.001
Albumin, g/dL	3744 (22.7)	2.8 (1.0)	1 (0.6)	3.4 (0.5)	<.001
Blood urea nitrogen, mg/dL	2818 (17.1)	19.0 (16.5)	0 (0.0)	13.5 (10.3)	<.001
Troponin I, ng/mL	12,026 (72.8)	0.06 (0.18)	135 (84.9)	0.05 (0.14)	.82
D-dimer, ng/mL	14,717 (89.1)	2311.8 (2646.5)	141 (88.7)	1125.5 (976)	.001
Lactate dehydrogenase, units/L	13,793 (83.5)	293.0 (259.8)	56 (35.2)	369.0 (188.0)	<.001
Ferritin, ng/mL	14,763 (89.4)	355.0 (819.2)	55 (34.6)	788.5 (875.8)	<.001
Procalcitonin, ng/mL	13,965 (84.6)	0.5 (1.4)	52 (32.7)	0.1 (0.2)	<.001
C-reactive protein, mg/dL	15,225 (92.2)	8.7 (13.7)	71 (44.7)	7.2 (10.2)	.30
Creatine kinase, units/L	13,668 (82.8)	76.0 (143.5)	93 (58.5)	81.5 (161.1)	.61
Sodium, mmol/L	2751 (16.7)	136.9 (4.8)	0 (0.0)	137.0 (4.3)	.56
Potassium, mmol/L	2750 (16.7)	4.0 (0.5)	0 (0.0)	4.0 (0.5)	.97
Creatinine, mg/dL	2757 (16.7)	1.0 (0.6)	0 (0.0)	0.8 (0.3)	<.001
Prothrombin time, s	6322 (38.3)	15.2 (3.2)	43 (27.0)	13.6 (1.5)	<.001
Oxygen saturation, %	7107 (43.0)	96.0 (2.5)	0 (0.0)	95.9 (2.4)	.10
Fraction of inspired oxygen, %	8125 (49.2)	28.5 (10.0)	35 (22.0)	27.7 (13.4)	.44

Model Performance: Predicting 48-Hour IMV

The AUROCs of all models for predicting 48-hour IMV risk are illustrated in Figure 2. Algorithm discrimination and other performance metrics for the COVID-19–like cohorts are presented for each model in Table S3 in Multimedia Appendix 1. At the prespecified threshold of 0.5, the XGBoost classifiers achieved the highest AUROCs (range 0.772-0.905) compared to other machine learning classifiers in each of the 16 training

cohorts for prediction of 48-hour risk for IMV. The best PPVs ranged between 0.583 and 0.767, and all models had an accuracy of 0.724 or higher and specificity of 0.786 or higher. The model trained with the influenza cohort was one of the worst-performing models, with an AUROC of 0.772 and PPV of 0.583 for IMV.

Figure 2. Receiver operating characteristic curves for model performance on predicting 48-hour invasive mechanical ventilation risk in the COVID-19 cohort, stratified by COVID-19–like cohorts. ARDS: acute respiratory distress syndrome; ROC: receiver operating characteristic.

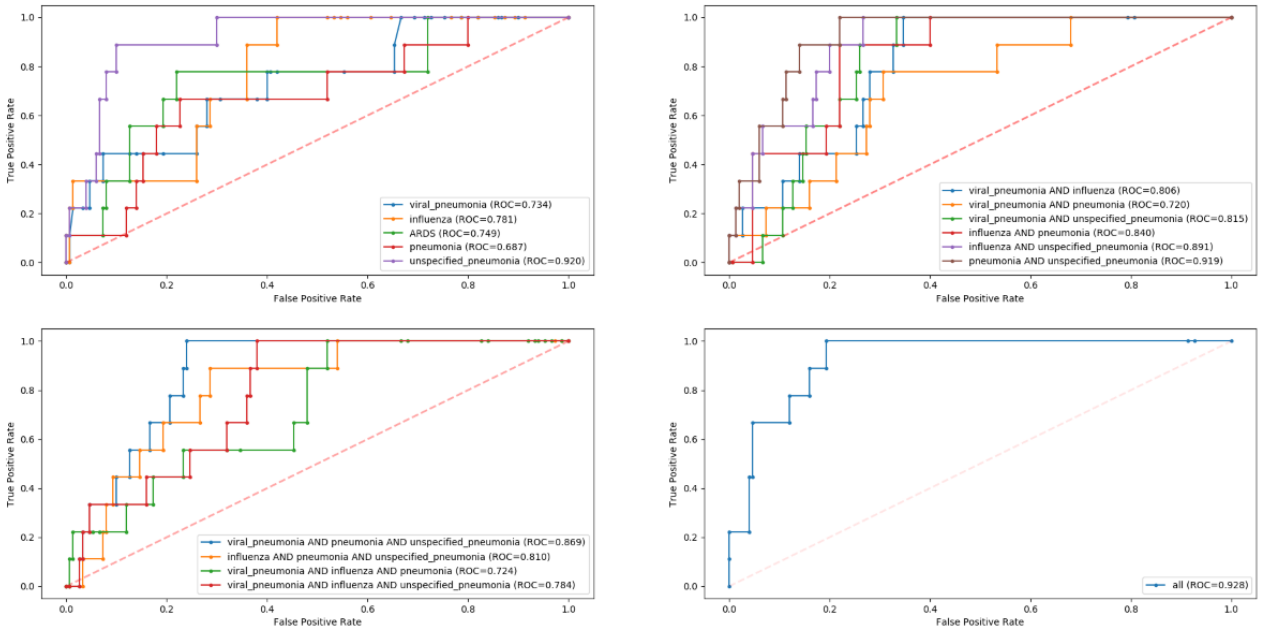


The pretrained models for 48-hour IMV prediction performed worse on the COVID-19 cohort than on the COVID-19–like cohorts. The model with the best performance on the new COVID-19 cohort was the XGBoost model trained on the viral pneumonia cohort (AUROC=0.826). The accuracy, precision, and sensitivity of the best model were 0.948, 0.423, and 0.289, respectively. Performance metrics for each model are presented in Table S4 in Multimedia Appendix 1. For negative cases at the patient-day level, the mean risk score was 0.09 (SD 0.124), with a minimum of 0.002 and a maximum of 0.859. For positive cases, the mean risk score was 0.31 (SD 0.23), with a minimum of 0.012 and a maximum of 0.823. At the patient level, the best-performing model was able to predict IMV in 48 hours for 7 of the 24 intubated patients with COVID-19. The model failed to predict the need for IMV in 48 hours for 17 patients (71%). Among the non-IMV patients, two were predicted to be at high risk of requiring IMV, although they were never intubated during hospitalization. Further details of the model hyperparameter optimization for predicting IMV can be found in Tables S5 and S6 in Multimedia Appendix 1.

Model Performance: Predicting All-Cause Mortality

Figure 3 and Table S7 in Multimedia Appendix 1 present detailed results of model performance for mortality prediction on the COVID-19–like cohorts. The XGBoost classifiers outperformed other machine learning classifiers in each of the 16 training cohorts for all-cause mortality prediction, with AUROC values ranging from 0.679 to 0.748.

Figure 3. Receiver operating characteristic curves for model performance on predicting inpatient mortality in the COVID-19 cohort, stratified by COVID-19–like cohorts. ARDS: acute respiratory distress syndrome; ROC: receiver operating characteristic.



Mortality prediction was better when testing on the COVID-19 cohort, with AUROC values ranging from 0.687 to 0.928. The best-performing model for the COVID-19 cohort was the XGBoost model trained on all 4 COVID-19–like cohorts (AUROC=0.928). The accuracy, precision, and sensitivity of the best model were 0.925, 0.286, and 0.222, respectively. The worst-performing model for the COVID-19 cohort was trained by using the pneumonia cohort, and achieved an AUROC of 0.687. Other results are presented in Table S8 in Multimedia Appendix 1. For all-cause mortality prediction at the patient level, the best-performing model predicted 22% of the deaths at the time of hospital admission. Further, our model predicted 5 deaths among the 109 patients discharged alive at the time of hospital admission.

Manual Chart Review

Manual chart review of 24 false positive and false negative cases was performed by author HC. Among the seven false positive cases for IMV prediction, one patient received IMV; however, this was not entered as structured data in the EHRs, therefore it was marked as a false positive for the model. Furthermore, two patients were extremely ill and close to receiving IMV. In the 17 false negative cases, 6 patients experienced rapid clinical deterioration in less than 24 hours or sometimes within 12 hours, and another 6 patients received scores that were close to the threshold. Additionally, our model incorrectly predicted the risk of death in 12 patients, including 5 false positive and 7 false negative cases. Based on the chart review, we confirmed that 3 of the 5 false positives were true false positives, while for the other two, the patients' conditions were severe and they were identified by the clinician as having a higher risk of death.

Algorithm Variable Importance

To identify the most salient features driving model prediction, we calculated SHA*P* values for the best-performing model. Among the top features, elevated fraction of inspired oxygen, total bilirubin, white blood cell count, lymphocyte count, D-dimer, and cardiac troponin, and lower albumin, oxygen saturation, and platelet count favored the classifier to predict an IMV event. Other important features from the best prediction model are in Figure 4. For mortality, being older and having higher blood urea nitrogen, potassium, and high-sensitivity cardiac troponin, and low albumin, oxygen saturation, and platelet count were the most influential factors in driving mortality prediction. Other important features associated with increased mortality risk included elevated total bilirubin and lower platelet count (Figure 5).

Figure 4. SHapley Additive exPlanations (SHAP) scores for identifying important features for prediction of invasive mechanical ventilation, including demographic information, vital values, and laboratory values. The color indicates whether the value of the feature is high (red) or low (blue).

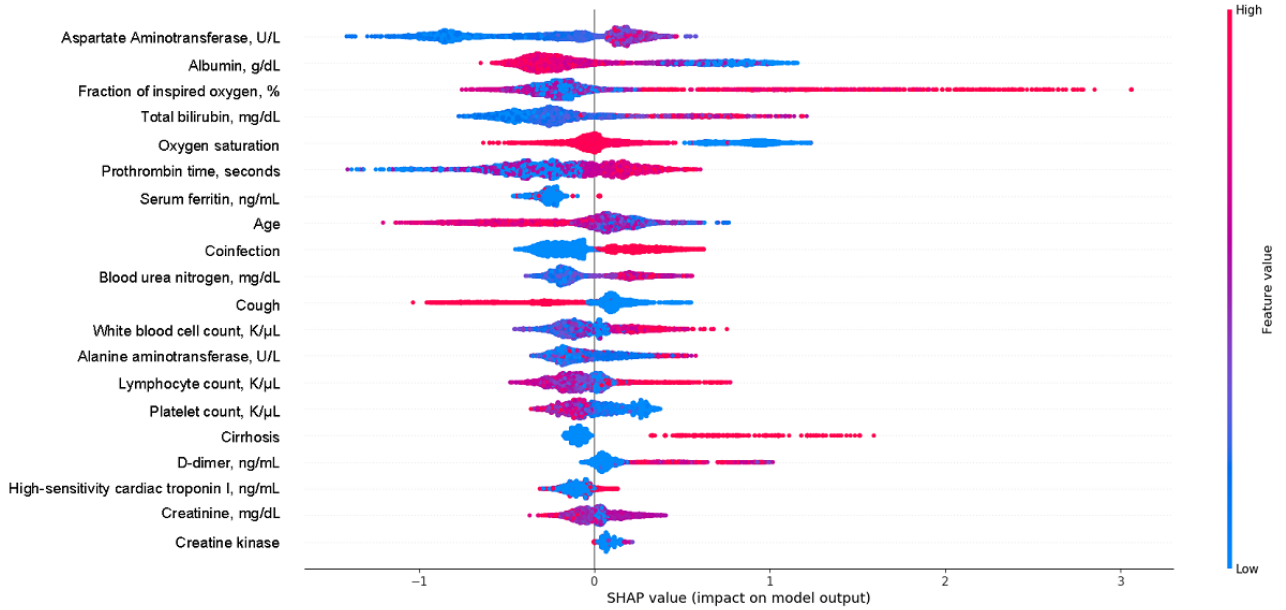
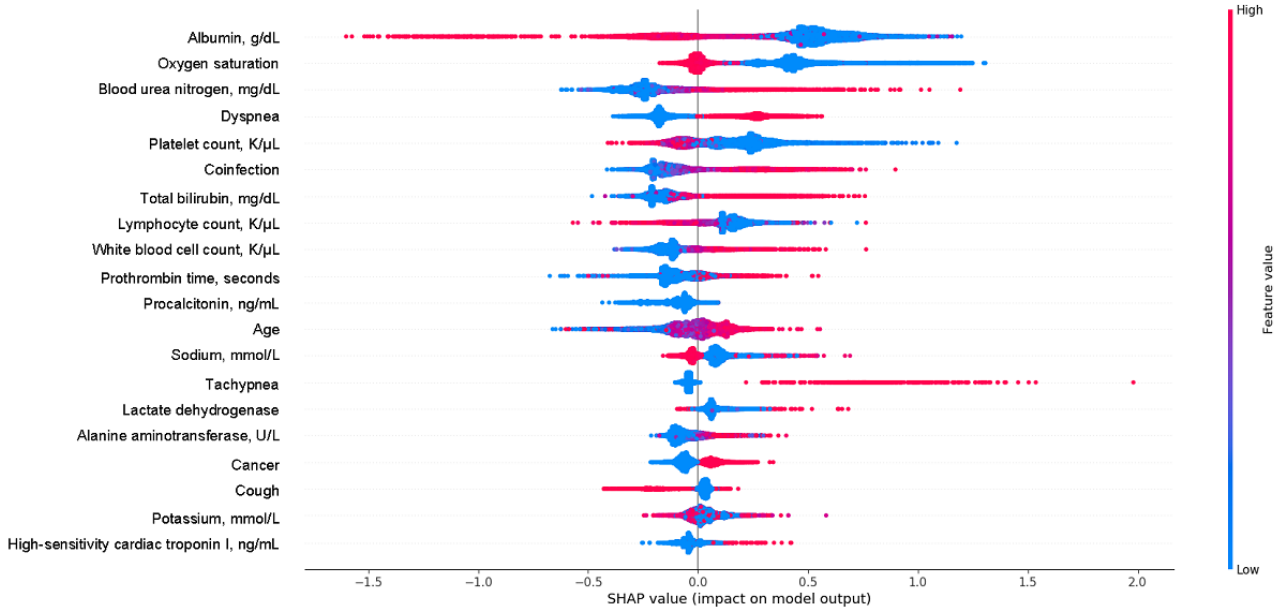


Figure 5. SHapley Additive exPlanations (SHAP) scores of the mortality prediction model. The color indicates whether the value of the feature is high (red) or low (blue).



Discussion

Principal Findings

In the clinical care of well-established diseases, literature and research are supplemented by the clinical judgment that is formed and refined through repeated episodes of care. Given the novelty of COVID-19, there is a lack of research evidence and clinical experience to inform clinical practice and guide care decisions. AI-enabled clinical decision support tools are promising to fill this gap and bolster clinical decision making. Early studies suggest that the disease manifestation, symptoms, and clinical course of COVID-19 resemble that of other respiratory infections, particularly pneumonia, influenza, and ARDS [5,30-32]. Due to the lack of robust, unbiased, representative data to train an AI model, we designed a framework to bootstrap existing retrospective data from COVID-19–like cohorts to predict IMV and all-cause mortality.

Our findings regarding the study cohort are consistent with previous international studies comparing patients with COVID-19 with patients previously hospitalized with other respiratory illnesses. The patients hospitalized with COVID-19 in our health system were younger and had fewer comorbidities than the COVID-19–like patients [33]. Patients with COVID-19 were less likely to have ever smoked compared to the COVID-19–like cohort. In addition, racial and ethnic minorities have been disproportionately affected by the disease, and we show similar trends in our health care system [32,34]. In addition to the variations in demographics and clinical outcomes, we

observed differences in symptoms at admission and laboratory values during the hospital stay. Importantly, significant differences were observed in IMV rates, time to IMV, and mortality rates between the two cohorts, with lower IMV and mortality rates in the patients with COVID-19 than in the COVID-19–like cohort, yet faster deterioration, as indicated by a shorter duration from admission to IMV.

The models, in general, performed well in the COVID-19–like cohorts, but less optimally in the COVID-19 cohort. Although the AUROC values of the best IMV and mortality prediction models in the COVID-19 cohort were good, the PPVs of both models were low, although the NPVs were high. Overall, the two models underestimated risk scores in patients with COVID-19. This can be explained by several factors. First, symptoms, laboratory values, and the proportions of missing data were different between COVID-19–like patients and patients with COVID-19, despite them sharing similar clinical manifestations and symptoms. Second, unlike the COVID-19–like cohorts, patients with COVID-19 had a broad spectrum of clinical manifestations, with critical courses that may involve fast deterioration and the need for IMV within 24 hours; thus, patients may have had limited signs of severe disease progression 48 hours prior to IMV. On the other hand, several clinical circumstances may affect clinical decisions. Clinicians may be reluctant to put patients on a ventilator due to the complexity of and complications associated with this invasive procedure. Patients or family members might also be hesitant to consent to the procedure out of fear of losing control [35]. Sometimes, patients have a Do Not Intubate code status, indicating they do not want to receive IMV in the event of a life-threatening situation. All these factors challenge the performance of our model, particularly the PPV. The model made a positive prediction for patients who require IMV, yet patients may not receive it in the end due to the factors mentioned above, and these are therefore considered false positives.

Traditional machine learning evaluation criteria, such as AUROC and PPV, were used to assess the performance of predicting the risk of IMV and all-cause mortality at the patient-day level; however, there is a lack of standard criteria to evaluate the model at the patient level when there are also multiple day-level data. When reporting the patient-level prediction results, a “strict” criterion was selected: the model correctly predicted the case only when the alarm occurred at 48 hours before IMV. If the prediction occurred too early or too close to IMV, it was considered wrong for that patient. If correct cases were determined as at least one alarm before IMV for those who were finally intubated and no alarm for non-IMV patients, the model performance would be greatly improved. The use of different standards for analyzing patient-level results can have clinical significance. For example, if a patient is predicted to be intubated during hospitalization, the provider needs to be reminded to pay more attention to avoid the rapid deterioration of the patient’s condition. Therefore, our framework provides important insight into the deterioration of patients with COVID-19 and the timing of that deterioration. Further studies are needed to explore evaluation criteria for this novel, emerging disease.

The performances of our COVID-19–like models suggest that the deterioration in the COVID-19 population in our health system is more similar to viral pneumonia than other respiratory illnesses. These results support our evolving understanding of the clinical characteristics of this disease state and support evidence that patients with COVID-19 are less like ARDS patients than was originally believed [36]. Although the scientific community has been shifting practices of treating the ventilation needs of patients with COVID-19 away from mimicking ARDS treatment, this work may be the first indication that there is a detectable demographic and pathophysiologic difference in the presenting characteristics of COVID-19 as well as the response to therapy.

This work has several clinical applications. It is notable that this model could make for an excellent screening tool for clinical deterioration in the inpatient setting because the model has a high positive likelihood ratio and high specificity. This indicates that a positive result truly indicates an increased probability of clinical respiratory decline for those admitted to hospital with COVID-19, even if the individual PPV for that “patient-day” is low. The implementation of this model would allow for enhanced monitoring of patients likely to require advanced respiratory support, especially during surge settings when there is a strain on staffing with advanced infectious disease or pulmonary training.

Although there is lower ability to predict patients in need of IMV within 48 hours or all-cause mortality at the “patient-day” level with great precision, this work can also be used to identify patients who are not at risk of clinical escalation. This means that this model can be used as a screening tool in our population, offering providers some confidence in the current level of care being appropriate rather than using valuable hospital resources on enhanced monitoring for patients who are less likely to need advanced medical management for respiratory failure. Such information can help with resource allocation and help providers triage patients who are less likely to become critically ill. Given the extreme stress and burden the COVID-19 pandemic has placed on the health care system, particularly on frontline workers, identifying patients who may need less focused attention may reduce some of the burden for health care systems that are already stretched thin. We anticipate that this work could be implemented across health care systems and therefore provide all codes and software needed to deploy these models. The external validation of our framework across systems could help in elucidating the clinical course of COVID-19 by refining the model in populations over time.

Limitations

While we envision many possible applications of our framework, we also recognize several limitations. First, our data come from a single health care system, and the results may not generalize to other health care systems that may have a different patient population or clinical practice. External validation would be required to reinforce our conclusions. Second, the sample sizes of the COVID-19–like and COVID-19 cohorts are different, which may bias our comparison in terms of demographics and clinical characteristics. However, despite this difference, these patients share similar clinical manifestations, histological

features, and disease progression. Third, our data set contained a relatively small number of deaths, and the model performance on the COVID-19 cohort could be unstable based on the limited number of patients. Future work is needed to validate the mortality prediction in other settings. Finally, the prediction of the risk of mortality used data from admission, limiting the performance of the model. Despite these limitations, important lessons have been learned from our experience of using pretrained machine learning models for disease severity prediction. It is feasible to pretrain a model using an unseen disease-like cohort but this requires special caution. First, selecting the most appropriate cohort, one which is similar to the clinical manifestation, pathological features, and disease progression of the targeted disease population, is essential for developing a successful machine learning model. Second, based on the nature of the disease and the type of data that are available, determining the right time frame for your machine learning model is crucial. We failed to predict whether or not a patient with COVID-19 would need IMV during their hospital stay using data obtained at hospital admission. Predictions may be hampered by the rapid deterioration seen in some patients with COVID-19 and changes in laboratory results.

Conclusions

In conclusion, our work demonstrates the feasibility of using existing data infrastructure and AI technology to guide critical care resource allocation in the early stages of a disease outbreak when not many cases have been observed and there is a lack of training data. Although the spread of COVID-19 has been exponential worldwide, most individual health care systems do not have a comprehensive, diverse, readily available data set of patients with COVID-19, which is necessary to develop, train, and validate essential AI models that may be used to guide clinical care. To date, many COVID-19–related AI models distributed through the scientific community have been trained and "validated" on only a handful of patients. However, given the lack of knowledge related to COVID-19, there is an urgent need to learn as much as possible about the disease, even if from small nonrepresentative populations. The framework we describe provides a strategy to mitigate this lack of data by identifying how and what we can learn from other COVID-19–like diseases. As we will likely deal with another wave of COVID-19 cases and other pandemics in the future, having a framework to rapidly design and train predictive models will have eminent value. Although using these COVID-19–like cohorts to learn about and predict outcomes of patients with COVID-19 may not be ideal, they provide an unbiased pathway to help guide clinical decisions when faced with this novel disease.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 64 KB - jmir_v23i2e23026_app1.docx]

References

1. Johns Hopkins Coronavirus Resource Center. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-07-28]
2. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72,314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
3. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, China Medical Treatment Expert Group for COVID-19. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA Intern Med 2020 Aug 01;180(8):1081-1089 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2033] [Medline: 32396163]
4. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. The Lancet Respiratory Medicine 2020 Apr;8(4):420-422. [doi: 10.1016/s2213-2600(20)30076-x]
5. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
6. Cummings MJ, Baldwin MR, Abrams D, Jacobson SD, Meyer BJ, Balough EM, et al. Epidemiology, clinical course, and outcomes of critically ill adults with COVID-19 in New York City: a prospective cohort study. Lancet 2020 Jun 06;395(10239):1763-1770 [FREE Full text] [doi: 10.1016/S0140-6736(20)31189-2] [Medline: 32442528]
7. Spellberg B, Haddix M, Lee R, Butler-Wu S, Holtom P, Yee H, et al. Community Prevalence of SARS-CoV-2 Among Patients With Influenzalike Illnesses Presenting to a Los Angeles Medical Center in March 2020. JAMA 2020 May 19;323(19):1966-1967 [FREE Full text] [doi: 10.1001/jama.2020.4958] [Medline: 32232421]
8. Grasselli G, Pesenti A, Cecconi M. Critical Care Utilization for the COVID-19 Outbreak in Lombardy, Italy: Early Experience and Forecast During an Emergency Response. JAMA 2020 Apr 28;323(16):1545-1546. [doi: 10.1001/jama.2020.4031] [Medline: 32167538]

9. Ranney ML, Griffeth V, Jha AK. Critical Supply Shortages - The Need for Ventilators and Personal Protective Equipment during the Covid-19 Pandemic. N Engl J Med 2020 Apr 30;382(18):e41. [doi: 10.1056/NEJMp2006141] [Medline: 32212516]
10. Rosenbaum L. Facing Covid-19 in Italy - Ethics, Logistics, and Therapeutics on the Epidemic's Front Line. N Engl J Med 2020 May 14;382(20):1873-1875. [doi: 10.1056/NEJMp2005492] [Medline: 32187459]
11. Auld S, Caridi-Scheible M, Blum J, Robichaux C, Kraft C, Jacob J, Emory COVID-19 Quality and Clinical Research Collaborative. ICU and Ventilator Mortality Among Critically Ill Adults With Coronavirus Disease 2019. Crit Care Med 2020 Sep;48(9):e799-e804 [FREE Full text] [doi: 10.1097/CCM.0000000000004457] [Medline: 32452888]
12. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020 May;8(5):475-481 [FREE Full text] [doi: 10.1016/S2213-2600(20)30079-5] [Medline: 32105632]
13. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
14. Guo Y, Liu Y, Lu J, Fan R, Zhang F, Yin X, et al. Development and validation of an early warning score (EWAS) for predicting clinical deterioration in patients with coronavirus disease 2019. MedRxiv Preprint published online on April 21, 2020. [doi: 10.1101/2020.04.17.20064691]
15. Ji D, Zhang D, Xu J, Chen Z, Yang T, Zhao P, et al. Prediction for Progression Risk in Patients With COVID-19 Pneumonia: The CALL Score. Clin Infect Dis 2020 Sep 12;71(6):1393-1399 [FREE Full text] [doi: 10.1093/cid/ciaa414] [Medline: 32271369]
16. Vaid A, Somani S, Russak A, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine Learning to Predict Mortality and Critical Events in a Cohort of Patients With COVID-19 in New York City: Model Development and Validation. J Med Internet Res 2020 Nov 06;22(11):e24018 [FREE Full text] [doi: 10.2196/24018] [Medline: 33027032]
17. Levy T, Richardson S, Coppa K, Barnaby DP, McGinn T, Becker LB, et al. Development and Validation of a Survival Calculator for Hospitalized Patients with COVID-19. MedRxiv Preprint published online on June 2, 2020. [FREE Full text] [doi: 10.1101/2020.04.22.20075416] [Medline: 32511640]
18. Zhang H, Shi T, Wu X, Zhang X, Wang K, Bean D, et al. Risk prediction for poor outcome and death in hospital in-patients with COVID-19: derivation in Wuhan, China and external validation in London. MedRxiv Preprint published online on May 3, 2020. [doi: 10.1101/2020.04.28.20082222]
19. Singh K, Valley T, Tang S, Li BY, Kamran F, Sjoding MW, et al. Validating a Widely Implemented Deterioration Index Model Among Hospitalized COVID-19 Patients. MedRxiv Preprint published online on April 29, 2020. [FREE Full text] [doi: 10.1101/2020.04.24.20079012] [Medline: 32511650]
20. Jewell NP, Lewnard JA, Jewell BL. Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections. JAMA 2020 May 19;323(19):1893-1894. [doi: 10.1001/jama.2020.6585] [Medline: 32297897]
21. Huang H, Cai S, Li Y, Li Y, Fan Y, Li L, et al. Prognostic Factors for COVID-19 Pneumonia Progression to Severe Symptoms Based on Earlier Clinical Features: A Retrospective Analysis. Front Med (Lausanne) 2020 Oct 5;7:557453 [FREE Full text] [doi: 10.3389/fmed.2020.557453] [Medline: 33123541]
22. Xie J, Hungerford D, Chen H, Abrams ST, Wang G, Wang Y, et al. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. MedRxiv Preprint published online on April 7, 2020. [doi: 10.1101/2020.03.28.20045997]
23. Sang S, Hernandez-Boussard T. Codes, models and synthesized data for paper 'Learning from past respiratory infections to predict COVID-19 Outcomes: A retrospective study'. 2020 Jul 01. URL: https://doi.org/10.25740/vg665fp4193 [accessed 2020-07-01]
24. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA 2020 May 12;323(18):1775-1776. [doi: 10.1001/jama.2020.4683] [Medline: 32203977]
25. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
26. Ma F, Yu L, Ye L, Yao DD, Zhuang W. Length-of-Stay Prediction for Pediatric Patients With Respiratory Diseases Using Decision Tree Methods. IEEE J Biomed Health Inform 2020 Sep;24(9):2651-2662. [doi: 10.1109/jbhi.2020.2973285]
27. Yang P, Wu T, Yu M, Chen F, Wang C, Yuan J, et al. A new method for identifying the acute respiratory distress syndrome disease based on noninvasive physiological parameters. PLoS One 2020 Feb 5;15(2):e0226962 [FREE Full text] [doi: 10.1371/journal.pone.0226962] [Medline: 32023257]
28. Bolourani S, Brenner M, Wang P, McGinn T, Hirsch JS, Barnaby D, Northwell COVID-19 Research Consortium. A Machine Learning Prediction Model of Respiratory Failure Within 48 Hours of Patient Admission for COVID-19: Model Development and Validation. J Med Internet Res 2021 Feb 10;23(2):e24246 [FREE Full text] [doi: 10.2196/24246] [Medline: 33476281]
29. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. arXiv Preprint posted online on May 22, 2017. [FREE Full text]

30. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA 2020 Mar 17;323(11):1061-1069 [FREE Full text] [doi: 10.1001/jama.2020.1585] [Medline: 32031570]
31. Zhu J, Ji P, Pang J, Zhong Z, Li H, He C, et al. Clinical characteristics of 3062 COVID-19 patients: A meta-analysis. J Med Virol 2020 Oct 24;92(10):1902-1914 [FREE Full text] [doi: 10.1002/jmv.25884] [Medline: 32293716]
32. Gold JAW, Wong KK, Szablewski CM, Patel PR, Rossow J, da Silva J, et al. Characteristics and Clinical Outcomes of Adult Patients Hospitalized with COVID-19 - Georgia, March 2020. MMWR Morb Mortal Wkly Rep 2020 May 08;69(18):545-550 [FREE Full text] [doi: 10.15585/mmwr.mm6918e1] [Medline: 32379729]
33. Burn E, You SC, Sena AG, Kostka K, Abedtash H, Abrahão MTF, et al. Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study. Nat Commun 2020 Oct 06;11(1):5009 [FREE Full text] [doi: 10.1038/s41467-020-18849-z] [Medline: 33024121]
34. Pan D, Sze S, Minhas JS, Bangash MN, Pareek N, Divall P, et al. The impact of ethnicity on clinical outcomes in COVID-19: A systematic review. EClinicalMedicine 2020 Jun;23:100404 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100404] [Medline: 32632416]
35. Macauley R. To trach or not to trach, that is the question. Paediatr Respir Rev 2019 Feb;29:9-13. [doi: 10.1016/j.prrv.2018.05.004] [Medline: 30665734]
36. Gattinoni L, Coppola S, Cressoni M, Busana M, Rossi S, Chiumello D. COVID-19 Does Not Lead to a 'Typical' Acute Respiratory Distress Syndrome. Am J Respir Crit Care Med 2020 May 15;201(10):1299-1300 [FREE Full text] [doi: 10.1164/rccm.202003-0817LE] [Medline: 32228035]

Abbreviations

AI: artificial intelligence
ARDS: acute respiratory distress syndrome
AUROC: area under the receiver operating characteristic curve
IMV: invasive mechanical ventilation
NPV: negative predictive value
PPV: positive predictive value
SMOTE: synthetic minority oversampling technique
SVM: support vector machine

Edited by C Basch; submitted 29.07.20; peer-reviewed by B Narayan, I Contreras, JA Benítez-Andrades; comments to author 18.09.20; revised version received 09.10.20; accepted 01.02.21; published 22.02.21.

Please cite as:
Sang S, Sun R, Coquet J, Carmichael H, Seto T, Hernandez-Boussard T
Learning From Past Respiratory Infections to Predict COVID-19 Outcomes: Retrospective Study
J Med Internet Res 2021;23(2):e23026
URL: https://www.jmir.org/2021/2/e23026
doi:10.2196/23026
PMID:33534724

Original Paper

Patients' Preferences for Artificial Intelligence Applications Versus Clinicians in Disease Diagnosis During the SARS-CoV-2 Pandemic in China: Discrete Choice Experiment

Taoran Liu[1,2*], BSc; Winghei Tsang[1*], MBBS; Fengqiu Huang[1], MPH; Oi Ying Lau[1], MBBS; Yanhui Chen[1], MBBS; Jie Sheng[1], BA; Yiwei Guo[3], BBA; Babatunde Akinwunmi[4,5], MPH, PhD, MM, MMSc; Casper JP Zhang[6], PhD, MPH; Wai-Kit Ming[1], MD, MPH, PhD, MMSc

[1]Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou, China

[2]Faculty of Economics and Business, University of Groningen, Groningen, Netherlands

[3]School of Finance and Business, Shanghai Normal University, Shanghai, China

[4]Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, MA, United States

[5]Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Harvard University, Boston, MA, United States

[6]School of Public Health, The University of Hong Kong, Hong Kong, China (Hong Kong)

[*]these authors contributed equally

Corresponding Author:
Wai-Kit Ming, MD, MPH, PhD, MMSc
Department of Public Health and Preventive Medicine
School of Medicine
Jinan University
West Huangpu Road 601
Guangzhou, 510000
China
Phone: 86 14715485116
Email: wkming@connect.hku.hk

Abstract

Background: Misdiagnosis, arbitrary charges, annoying queues, and clinic waiting times among others are long-standing phenomena in the medical industry across the world. These factors can contribute to patient anxiety about misdiagnosis by clinicians. However, with the increasing growth in use of big data in biomedical and health care communities, the performance of artificial intelligence (Al) techniques of diagnosis is improving and can help avoid medical practice errors, including under the current circumstance of COVID-19.

Objective: This study aims to visualize and measure patients' heterogeneous preferences from various angles of AI diagnosis versus clinicians in the context of the COVID-19 epidemic in China. We also aim to illustrate the different decision-making factors of the latent class of a discrete choice experiment (DCE) and prospects for the application of AI techniques in judgment and management during the pandemic of SARS-CoV-2 and in the future.

Methods: A DCE approach was the main analysis method applied in this paper. Attributes from different dimensions were hypothesized: diagnostic method, outpatient waiting time, diagnosis time, accuracy, follow-up after diagnosis, and diagnostic expense. After that, a questionnaire is formed. With collected data from the DCE questionnaire, we apply Sawtooth software to construct a generalized multinomial logit (GMNL) model, mixed logit model, and latent class model with the data sets. Moreover, we calculate the variables' coefficients, standard error, *P* value, and odds ratio (OR) and form a utility report to present the importance and weighted percentage of attributes.

Results: A total of 55.8% of the respondents (428 out of 767) opted for AI diagnosis regardless of the description of the clinicians. In the GMNL model, we found that people prefer the 100% accuracy level the most (OR 4.548, 95% CI 4.048-5.110, *P*<.001). For the latent class model, the most acceptable model consists of 3 latent classes of respondents. The attributes with the most substantial effects and highest percentage weights are the accuracy (39.29% in general) and expense of diagnosis (21.69% in general), especially the preferences for the diagnosis "accuracy" attribute, which is constant across classes. For class 1 and class 3, people prefer the AI + clinicians method (class 1: OR 1.247, 95% CI 1.036-1.463, *P*<.001; class 3: OR 1.958, 95% CI

1.769-2.167, *P*<.001). For class 2, people prefer the AI method (OR 1.546, 95% CI 0.883-2.707, *P*=.37). The OR of levels of attributes increases with the increase of accuracy across all classes.

Conclusions: Latent class analysis was prominent and useful in quantifying preferences for attributes of diagnosis choice. People's preferences for the "accuracy" and "diagnostic expenses" attributes are palpable. AI will have a potential market. However, accuracy and diagnosis expenses need to be taken into consideration.

(J Med Internet Res 2021;23(2):e22841) doi:10.2196/22841

KEYWORDS

discrete choice experiment; artificial intelligence; patient preference; multinomial logit analysis; questionnaire; latent-class conditional logit; app; human clinicians; diagnosis; COVID-19; China

Introduction

The phenomenon of uneven allocation and distribution of high-quality doctor resources has existed for centuries with the history of the modern medical industry, which brings about a series of problems such as gaps in diagnosis accuracy, speed, or accessibility in rural areas. According to a recent study, over 12 million patients in the United States have experienced one or more misdiagnoses, and the misdiagnosis rate is 5.08% [1]. Some developing countries still face the problem of scarcity of doctors. The World Health Organization (WHO) has suggested that 2.5 doctors per 1000 people are needed to guarantee primary health care [2]. However, it has been reported that there were only 1.9 doctors per 1000 people in 2017 in China, and 45% of WHO member countries still have less than 1 doctor per 1000 people [3]. Therefore, new medical technology, such as artificial intelligence (AI) technologies, urgently needs to be improved.

The recent outbreak of an epidemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a severe threat to public health. The COVID-19 pandemic, with lockdowns and unprecedented restrictions on movement, brings a newer angle to the importance of AI diagnosis. With big data growth in biomedical and health care communities, AI is increasingly applied to antiepidemic medical practice.

Medical AI can be classified into eight main fields: medical imaging and diagnosis, medical research, medical risk analysis, drug mining, virtual nurse assistant, prognostics and health management, mental health, and nutrition [4,5]. AI diagnosis and treatment technologies have become increasingly mature and are expected to become mainstream. As far as we know, there is no clear proof of how health outcomes or costs interdepend and correlate [6-8], although AI diagnosis was assumed certainly to be the higher cost-performance option from a medical angle. Such a lack of proof of clear superiority of either diagnosis method gives prominence to the patients' and medical institutions' preferences. Thus, the acceptability may be enhanced by choosing and adapting a diagnostic program to cater to patients' preferences.

In addition, the Ministry of Industry and Information Technology of China demanded an increase in the praxis of artificial intelligence in the precise prevention and control of epidemics [9]. AI algorithms combine chest computed tomography imaging reports with clinical symptoms, medical history, and laboratory examination to rapidly diagnose patients as infected by SARS-CoV-2. The AI system slightly reduced misdiagnosis by radiologists [10]. The COVID-19 detection neural network, a deep learning model, can precisely detect SARS-CoV-2 and distinguish it from other pneumonia [11]. Artificial neural network modeling of SARS-CoV-2 morbidity across the United States illustrated that a single-hidden-layer multilayer perceptron could interpret nearly 65% of the correlation with ground truth for the prognostication [12].

A few studies [13,14] have focused on the effect of outpatient waiting time, diagnosis time, follow-up after diagnosis, etc in patients' decision making and concluded that these factors play a vital role in patients' trade-off and relative policy making. However, with the development of AI in medicine and rise of AI diagnosis, patients start shifting their focus to the accuracy and expense of AI diagnosis. Thus, we aim to fill the gap that exists because almost no studies focus on the effects of accuracy or other attributes of AI diagnosis and clinicians in patients' choice.

The objectives of this paper are to measure the extent of patients' preferences for a range of characteristics of an AI diagnosis scheme in China and to determine what characteristics are more attractive and make AI a better alternative to defeat traditional medical methods. A technique that is in common use for visualizing preferences is the discrete choice experiment (DCE), in which different alternatives with various attributes are given in the form of a questionnaire to people who are invited to choose options. In this paper, we will construct and make a comparison of these three models: mixed logit (MXL) model, generalized multinomial logit (GMNL) model, and latent class model (LCM). Moreover, the importance of attribute levels and preference heterogeneity must be compared and reported when people are considering any AI diagnosis service.

Methods

Overview

We designed multiple choices in different scenarios consisting of 6 different randomly selected attributes using conjoint-related techniques. Questionnaires were created by Sawtooth software's Lighthouse Studio modules (version 9.8.1) for general interviews and choice-based conjoint (CBC) scenario design. Respondents were aged 18 to 85 years. Meanwhile, in this data analysis section, we aimed to visualize and measure the percentage weight and importance of different attributes with selected models. From the perspective of public health, McFadden's conditional logit [15], which is also known as multinomial logit (MNL) [16], is widely applied to organize, analyze, and predict

the data we had and help with further analysis of statistical significance. However, the foundation of using this model is that we acquiesce that unobserved heterogeneity of preference does not exist across respondents. Therefore, we then need to introduce a MXL model [17] and a GMNL model [18], both of which take unobserved heterogeneity of preference into account. The LCM [19] might also be an appropriate model to apply here, since it divides the data into various groups with fixed segmented size and the evident probability of latent membership [20]. With different latent classes, we can clearly distinguish the most important attributes or attributes' levels of each class and summarize these attributes with a significant percentage weight.

Principle of DCE

Random utility theory [21] is the basic principle of DCE. The principle assumes that all the choice selectors have *M* different choice alternatives, and each choice alternative corresponds to a utility *W*. The utility *W* is consistent with a combination of fixed and random utility. The stationary utility *U* can be explained by some observable elements *x*, while the random factor ε represents the influence and interference unobserved utility and possible error. The goal of choice selectors is to choose the best combination with the supreme utility; then the probability of each combination alternative being selected can be expressed as a function of its fixed utility: $P = F(U)$. The specific form of the function depends on the distribution of random effects. In most model settings, the utility that is optical, U_v, will be expressed as a linear combination of elements *x*, that is, $U_v = \beta x$. β is a coefficient, and its value and significance level can be estimated from the observation data.

Selection of Attributes

Based on the relevant literature [22-24], we have assumed that the patients' preference or satisfaction with the medical choices mainly depends on some specific features that make up the essential attributes of our experiment. Moreover, we performed a pilot test to get the attributes and levels that we need in our research. Patients in the outpatient queue of the First Affiliated Hospital of Jinan University (Guangzhou Overseas Chinese Hospital) and the First Affiliated Hospital of Sun Yat-sen University were interviewed and invited to have a discussion of what attributes patients are most likely to attach importance to. In addition, the possible attributes' levels are hypothesized and set in a certain sequence in our questionnaire; for instance, accuracy ranks from 0% to 100%.

Therefore, in our questionnaire, six attributes have been contained for our experiment: (1) diagnostic method; (2) outpatient waiting time before being asked; (3) diagnosis time; (4) accuracy (ratio of correct diagnosis); (5) follow-up after diagnosis (whether the outpatient doctor/AI doctor can follow up and follow up at any time); and (6) diagnostic expenses. Every attribute and its levels are presented in Table 1.

Table 1. 6 different attributes hypothesized and their levels in discrete choice experiment questionnaire.

Diagnosis methods	Levels
Diagnostic methods	Clinicians' diagnosis; AI diagnosis + clinicians' confirmation; AI diagnosis
Outpatient waiting time	0 min; 20 min; 40 min; 60 min; 80 min; 100 min
Diagnosis time	0 min; 15 min; 30 min
Accuracy	60%; 70%; 80%; 90%; 100%
Follow-up after diagnosis	Yes; No
Diagnostic expenses[a]	¥0; ¥50; ¥100; ¥150; ¥200; ¥250

[a]A currency exchange rate of ¥1=US $0.15 is applicable.

Questionnaire and DCE Design

The questionnaire contains two sections. In the first section, which is also known as demographic questions, we aim to allocate the respondents' basic information: age, gender, and highest education level. In the second section, we use the CBC function in Sawtooth software to create various combinations of scenarios for respondents to choose from.

When we use the factorials method [25] to analyze the attributes to give the combinations of scenarios, we encounter several obstacles. Since we have 6 attributes that give 3240 (3×6×3×5×2×6) possibilities, we assume that we have 6 random questions and 100 sets, which also gives 600 different combinations. The difficulty is how to extract 600 representative combinations from 3240 combinations and obey two basic principles [26] at the same time: (1) balance and (2) orthogonality. Balance means each attribute level appears equally often within an attribute, and level 1 in attribute 1 equals level 2 in attribute 2. Orthogonality means each pair of levels appears equally often across all pairs of attributes. However, it is almost unrealistic to handle such a huge task. Therefore, we use Sawtooth software to help us select suitable combinations. We set six random tasks, one fixed task, and two concepts per task (excluding "None Option"). Meanwhile, we set the sample size 500 and assume that 5% of respondents would choose "None Option" and finally design the test.

The standard error of almost all the attributes' levels is <0.05. Since the expense is a continuous variable of which the standard error could be slightly higher than 0.05, a sample size of 500 should be sufficient for our experiment. One of the CBC tasks of the DCE questionnaire has been presented in Table 2.

Table 2. An example scenario of choice-based conjoint in the questionnaire.

Attributes	Doctor A	Doctor B	None
Diagnostic methods	AI diagnosis	Clinicians' diagnosis	
Outpatient waiting time (min)	0	20	
Accuracy (%)	80	60	
Follow-up after diagnosis	No	Yes	
Diagnostic expenses (¥)[a]	150	200	
Which method would you choose?	Choose Doctor A	Choose Doctor B	No choice

[a]A currency exchange rate of ¥1=US $0.15 is applicable.

Data Collection Procedure

We sent our website link containing our DCE questionnaire through social media apps such as Facebook and WeChat. In addition, we gave respondents some rewards such as Mi bands or a small cash payment if the respondents could fully complete the questionnaire.

Statistical Analysis

GMNL Model and MXL Model

With collected data, some models can be applied in our analysis; first is the GMNL model. We do not use the MXL model or the scaled-multinomial logit (S-MNL) model here since the GMNL model, which is developed by Fiebig (2010) [27], nested a MXL model and scaled multinomial model. Meanwhile, the GMNL model could accurately describe consumers' preferences and heterogeneity. According to Fiebig et al, the probability of respondent i choosing alternative j in choice situation t is given as follows:

Here, β_i is a vector of an individual-specific parameter and the individual coefficients of independent variables can be described as follows:

$$\beta i = \sigma i\beta + \{\gamma + \sigma i(1-\gamma)\}\eta i$$

where the β is a constant vector, the "effect" σ_i in our model is a parameter of individual-specific scale, and γ is a parameter that decides how σ_i and η_i are different in some degrees. With this equation, when σ_i equals 1, our GMNL model will become an MXL model. Meanwhile, when the variance of η_i becomes 0, our GMNL model will turn into a S-MNL model. When σ_i equals 1 and variance of η_i equals 0 are both satisfied, the GMNL model will transform into an MNL model. Sawtooth will be needed to help run the coefficients of all attributes, standard errors, and *t* ratios to calculate *P* values. Differences between attributes are also needed to calculate the odds ratios. We calculate the odds ratio using the following equation:

$$odds\ ratio = exp\ (current\ effect - reference\ effect)$$

LCM

In addition, the LCM will be applied. LCM is a latent variable model because the latent variable is discrete. According to Greene and Hensher (2003) [28], the principle of LCM is that the observable attributes and latent heterogeneity decide the individual behavior. The heterogeneity changes with the unobserved factors. This model is used to sort individuals into a set of classes with a certain segmented size and scale, and different effects of each class have been estimated for different attributes. As well, the LCM will help us measure the differences and similarities of preference across classes of respondents. Attributes' importance and part worth utilities will also be needed for visual analysis and comparison of attributes and deciding which attribute is the most essential from the people's perspective. Additionally, the average maximum membership probability will help predict the certainty of the class into which respondents are divided.

Results

Respondents

428 participants (aged 18-85 years) who provided complete data were included in the analysis. Among those with complete data, 206 (48.1%) were male and 222 (51.9%) were female, while 2 of them were pregnant women.

Attributes' Levels and Utility Report

The average utility values of all the attributes' levels were measured using the Utility Scaling Method with zero-centered differences. The highest utility levels of the six hypothesized attributes are "AI diagnosis + clinicians' confirmation" (of "diagnosis method"), "20 min" (of "outpatient waiting time"), "15 min" (of "diagnosis time"), "100%" (of "accuracy"), "Yes" (of "follow-up"), and "¥0" (of "diagnostic expenses"), respectively. The most important attribute is "accuracy," meaning that the majority of respondents ranked that attribute (Table 3).

Table 3. The utility report of different attributes' levels.

Attributes and levels	Utility
Diagnostic methods	
Clinician	−11.51
AI + clinician	57.64
AI	−46.13
Outpatient waiting time (min)	
0	12.57
20	35.41
40	4.41
60	−27.99
80	−24.40
Diagnosis time (min)	
0	−7.02
15	4.14
30	2.88
Accuracy (%)	
60	−116.31
70	−60.65
80	−2.24
90	59.75
100	119.44
Follow-up after diagnosis	
Yes	27.88
No	−27.88
Diagnostic expenses (¥)[a]	
0	47.91
50	32.93
100	32.25
150	−5.92
200	−24.91
250	−82.25
None	
N/A[b]	−235.59

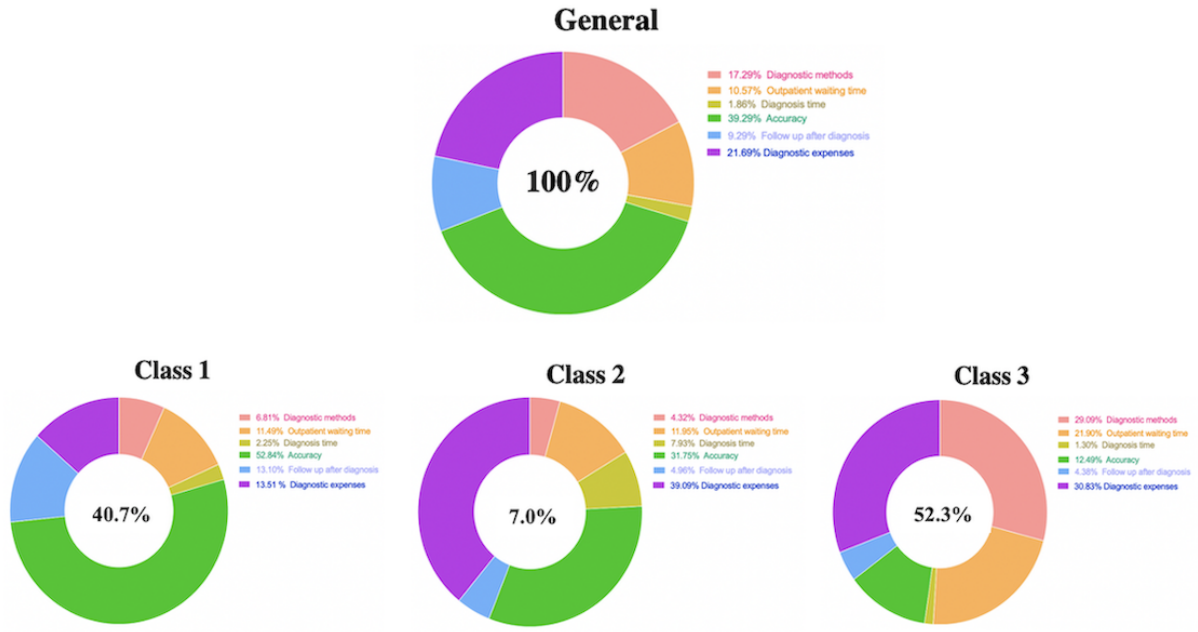
[a]A currency exchange rate of ¥1=US $0.15 is applicable.

[b]N/A: not applicable.

Logit Result of DCE and Attributes' Percentage Importance

In general, it is obvious that attribute "accuracy" was the most important factor and most preferred when facing the diagnosis. As is presented in Figure 1, the percentage importance of attribute "accuracy" was 39.29%, which undoubtedly shows the position of accuracy of diagnosis in people's minds. Attributes "diagnostic expenses" and "diagnostic method" ranked second and third, respectively.

Figure 1. Percentage importance of attributes in general condition and latent class condition.



The results of the logit analysis of all attributes' levels are presented in Table 4. For the diagnosis method, the coefficient of level "AI + clinician" is positive, which means that the level "AI + clinician" is positively correlated with people's preference and utility. In addition, the coefficients of levels "0 min," "20 min," and "40 min" (of "outpatient waiting time") are positive, and others are negative. It is clear that people have a preference for shorter outpatient waiting time. For other attributes' levels, people prefer to choose the higher accuracy level, greater possibility to follow up after diagnosis, and lower diagnosis expense. However, for the attribute "diagnosis time," people unexpectedly prefer a longer diagnosis time rather than the "0 min" option.

Table 4. The result of logit analysis of preference in general (N=428).

Attributes and levels	Coefficient	Standard Error	*P* value	Odds ratio	95% CI
Diagnostic methods					
Clinician	−0.07226	0.03829	.06	Reference	
AI + clinician	0.37192	0.0386	<.001	1.559	(1.446-1.682)
AI	−0.29966	0.03926	<.001	0.797	(0.738-0.860)
Outpatient waiting time (min)					
0	0.09507	0.05818	.10	Reference	
20	0.22298	0.05815	<.001	1.136	(1.014-1.274)
40	0.03253	0.05965	.59	0.939	(0.836-1.056)
60	−0.18624	0.0586	.002	0.755	(0.673-0.847)
80	−0.16435	0.05915	.006	0.771	(0.687-0.866)
Diagnosis time (min)					
0	−0.04446	0.03857	.25	Reference	
15	0.03001	0.03853	.44	1.077	(0.999-1.162)
30	0.01444	0.03831	.71	1.061	(0.984-1.143)
Accuracy (%)					
60	−0.74382	0.06294	<.001	Reference	
70	−0.4016	0.05957	<.001	1.408	(1.253-1.582)
80	−0.0097	0.057	.87	2.084	(1.863-2.330)
90	0.38431	0.05787	<.001	3.090	(2.759-3.461)
100	0.77081	0.05943	<.001	4.548	(4.048-5.110)
Follow-up after diagnosis					
Yes	0.18169	0.02415	<.001		
No	−0.18169	0.02415	<.001		
Diagnostic expenses (¥)[a]					
0	0.30678	0.06632	<.001	Reference	
50	0.22572	0.06606	<.001	0.922	(0.810-1.050)
100	0.20776	0.06673	.002	0.906	(0.795-1.032)
150	−0.04055	0.06692	.55	0.707	(0.620-0.806)
200	−0.16992	0.06693	.01	0.621	(0.545-0.708)
250	−0.52978	0.06916	<.001	0.433	(0.378-0.496)

[a]A currency exchange rate of ¥1=US $0.15 is applicable.

For the *P* value of these attributes' levels, we assume that if the *P* value of a level is less than .05, then this level is statistically significant; when the *P* value of a level is less than .001, then this level is extremely statistically significant. We found that "AI + clinician" and "AI diagnosis" for "diagnosis methods" are extremely statistically significant; "20 min" of outpatient waiting time is extremely statistically significant; "60 min" and "80 min" are statistically significant; "60%," "70%," "90%," "100%" of "accuracy" are extremely statistically significant; both levels of attribute "follow up after diagnosis" are extremely statistically significant. "¥0'" of attribute "diagnostic expenses" is extremely statistically significant, and all the other levels of diagnostic expenses are statistically significant.

The odds ratio is a commonly used indicator in case-control epidemiological studies. In our analysis and calculation results (Table 4), we find that some odds ratios of the attributes' levels compared to the reference level are greater than one, which means that the probability of people's choosing of these levels is higher than the previous one or the reference. Taking the level "Clinician" of attribute "diagnostic methods" as the reference, level "AI + clinician" has an odds ratio of 1.559 (95% CI 1.446-1.682). The odds ratio of level "20 min" (of "outpatient waiting time") is 1.136 (95% CI 1.014-1.274) with the reference of level "0 min." Levels "15 min" and "30 min" (of "diagnosis time") have odds ratios 1.077 and 1.067, respectively, with the reference of level "0 min" (95% CI 0.999-1.162 and 0.984-1.143, respectively). All the levels of the attribute

"accuracy" were greater than one, which means the preference weights increase with the accuracy. Meanwhile, all the odds ratios of expense compared to the reference are smaller than zero, which refers to a preference of "free diagnosis" for the majority of people.

Latent Class Analysis Result

We compared these potential models and selected the model that maximized the area under the receiver operating characteristic curve and minimized the Akaike information criterion (AIC) [29,30] and Bayesian information criterion (BIC) [31] to penalize for model complexity. According to AIC, 5 classes should be the best choice for our model. However, if we must choose with BIC, then a 2-class option should be the most appropriate one, since 2-class has the lowest BIC. Under such circumstances, we compare ABIC, which means sample size–adjusted BIC [32] and involves sample size value. After comparing, the 3-class option has the lowest value of ABIC. Therefore, the most suitable number of latent classes in our model was 3 (Table 5 and 6). First of all, we divided all 428 respondents into 3 classes with segment sizes of 174 (40.7%), 30 (7.0%), and 224 (52.3%). The average maximum membership probability is around 0.87 and the percent certainty was 35.30, which is relatively low, meaning that there was not much uncertainty according to the respondents divided into classes.

For class 1, *t* ratios for attributes "diagnostic expenses," "'accuracy," and "follow-up after diagnosis" were significant across all treatment modalities. Attribute "accuracy" was the most important factor for patients with 52.8%, followed by "diagnostic expenses" and "follow-up after diagnosis" with percentage importance 13.51% and 13.10%, respectively (Figure 1). Meanwhile, the span of percentage weights of "accuracy" is obvious (Figure 2), from −2.33 to 2.52. Meanwhile, the span of attribute "diagnostic expenses" is from −0.819 to 0.423. The preference weights of "diagnostic expenses" decrease with the increasing of the expenses. The preference weight of attribute "follow-up after diagnosis" is from −0.602 to 0.602. This is −0.602 for "No follow-up" and 0.602 for "follow-up", which is symmetrical. In addition, the odds ratio of level "AI + clinicians" was 1.247 (95% CI 1.036-1.463), meaning that the majority of patients prefer durable treatments over a single treatment. As well, all the levels of attribute "diagnosis time" with the reference level "0 min" are larger than one. At the same time, the odds ratio (Table 7) of the levels of attribute "accuracy" increases with the accuracy rate, meaning that people's preference weight increases with the accuracy.

Table 5. Result of 3 latent classes' conditional logit analysis.

Attributes and levels	Class 1, n=174 (40.7%)			Class 2, n=30 (7.0%)			Class 3, n=224 (52.3%)		
	Coefficient	SE	*P* value	Coefficient	SE	*P* value	Coefficient	SE	*P* value
Diagnostic methods									
Clinician	0.062	0.083	.046	−0.175	0.282	.54	−0.153	0.050	.003
AI + clinician	0.282	0.082	<.001	−0.085	0.287	.77	0.518	0.052	<.001
AI	−0.344	0.085	<.001	0.260	0.286	.37	−0.365	0.051	<.001
Outpatient waiting time (min)									
0	0.530	0.128	<.001	−0.385	0.476	.43	−0.028	0.077	.72
20	0.147	0.124	.24	−0.128	0.422	.76	0.351	0.079	<.001
40	−0.057	0.132	.67	0.819	0.365	.03	0.091	0.080	.25
60	−0.095	0.127	.45	−0.330	0.455	.48	−0.314	0.079	<.001
80	−0.526	0.122	<.001	0.023	0.414	.96	−0.100	0.079	.20
Diagnosis time (min)									
0	−0.117	0.084	.17	0.317	0.280	.27	−0.021	0.051	.68
15	0.089	0.082	.28	−0.481	0.345	.17	0.018	0.051	.72
30	0.028	0.082	.73	0.164	0.280	.56	0.003	0.051	.95
Accuracy (%)									
60	−2.337	0.166	<.001	−1.717	0.803	.04	−0.209	0.080	.01
70	−1.170	0.131	<.001	−0.693	0.555	.22	−0.129	0.078	.10
80	−0.050	0.112	.65	0.353	0.457	.45	0.034	0.078	.67
90	1.036	0.122	<.001	0.577	0.406	.17	0.135	0.079	.09
100	2.522	0.169	<.001	1.480	0.370	<.001	0.170	0.079	.03
Follow-up after diagnosis									
Yes	0.603	0.059	.003	0.250	0.207	.24	0.066	0.031	.04
No	−0.603	0.059	.03	−0.250	0.207	.24	−0.066	0.031	.035
Diagnostic expenses (¥)[a]									
0	0.424	0.140	.003	0.313	0.492	.53	0.398	0.090	<.001
50	0.324	0.143	.03	1.831	0.412	<.001	0.131	0.088	.14
100	0.289	0.142	.04	−0.228	0.506	.66	0.236	0.090	.009
150	0.144	0.149	.33	0.284	0.460	.54	−0.102	0.089	.25
200	−0.361	0.146	.01	−2.106	0.983	.04	−0.123	0.089	.17
250	−0.819	0.144	<.001	−0.093	0.533	.86	−0.538	0.091	<.001

[a]A currency exchange rate of ¥1=US $0.15 is applicable.

Table 6. Percent certainty and information criteria for model with 3 latent classes.

Characteristic	Value
Certainty (%)	35.307
Akaike information criterion	3580.631
Bayesian information criterion	3922.719
Sample size–adjusted Bayesian information criterion	3735.263

Figure 2. Latent class percentage weights in class 1. AI: artificial intelligence; RMB: yuan renminbi.

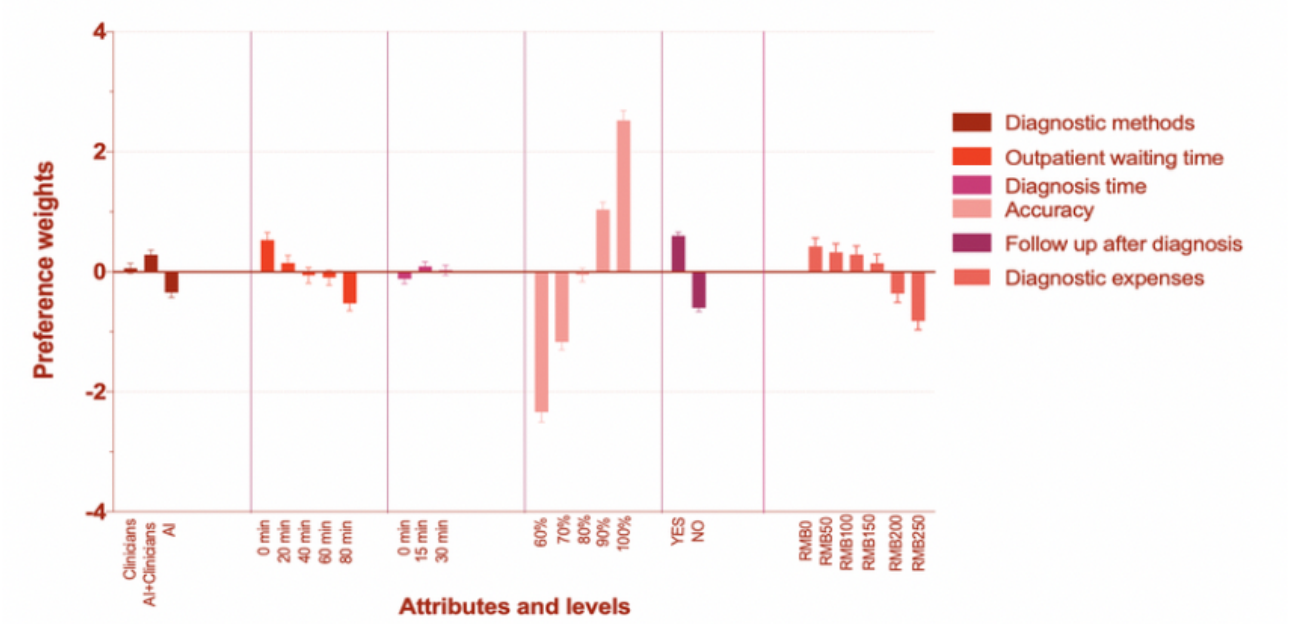


Table 7. The odds ratios and confidence intervals of attributes' levels in 3 classes.

Attributes and levels	Class 1, n=174 (40.7%)		Class 2, n=30 (7.0%)		Class 3, n=224 (52.3%)	
	OR[a]	95% CI	OR	95% CI	OR	95% CI
Diagnostic methods						
Clinician	Reference	N/A[b]	Reference	N/A	Reference	N/A
AI + clinician	1.247	(1.036-1.463)	1.094	(0.624-1.920)	1.958	(1.769-2.167)
AI	0.666	(0.564-0.787)	1.546	(0.883-2.707)	0.809	(0.732-0.895)
Outpatient waiting time (min)						
0	Reference	N/A	Reference	N/A	Reference	N/A
20	0.681	(0.535-0.868)	1.293	(0.566-2.957)	1.460	(1.252-1.703)
40	0.556	(0.429-0.720)	3.332	(1.628-6.821)	1.126	(0.963-1.316)
60	0.535	(0.418-0.686)	1.057	(0.433-2.580)	0.751	(0.643-0.877)
80	0.348	(0.274-0.442)	1.504	(0.688-3.388)	0.930	(0.797-1.085)
Diagnosis time (min)						
0	Reference	N/A	Reference	N/A	Reference	N/A
15	1.229	(1.047-1.444)	0.450	(2.229-0.885)	1.040	(0.942-1.149)
30	1.156	(0.986-1.357)	0.858	(0.469-1.485)	1.024	(0.927-1.132)
Accuracy (%)						
60	Reference	N/A	Reference	N/A	Reference	N/A
70	3.214	(2.484-4.159)	2.785	(0.938-8.271)	1.084	(0.930-1.263)
80	9.849	(7.912-12.258)	7.931	(3.240-19.417)	1.275	(1.095-1.485)
90	29.173	(22.962-37.064)	9.920	(4.480-21.962)	1.411	(1.207-1.648)
Follow-up after diagnosis						
Yes	Reference	N/A	Reference	N/A	Reference	N/A
No	0.300	(0.267-0.337)	0.607	(0.405-0.910)	0.876	(0.824-0.931)
Diagnostic expenses (¥)[c]						
0	Reference	N/A	Reference	N/A	Reference	N/A
50	0.905	(0.683-1.199)	4.563	(2.037-10.222)	0.766	(0.644-0.911)
100	0.847	(0.662-1.154)	0.583	(0.216-1.571)	0.851	(0.713-1.015)
150	0.756	(0.565-1.102)	0.972	(0.394-2.394)	0.606	(0.509-0.722)
200	0.456	(0.343-0.607)	0.089	(0.013-0.612)	0.594	(0.499-0.707)
250	0.289	(0.217-0.383)	0.666	(0.234-1.895)	0.392	(0.328-0.469)

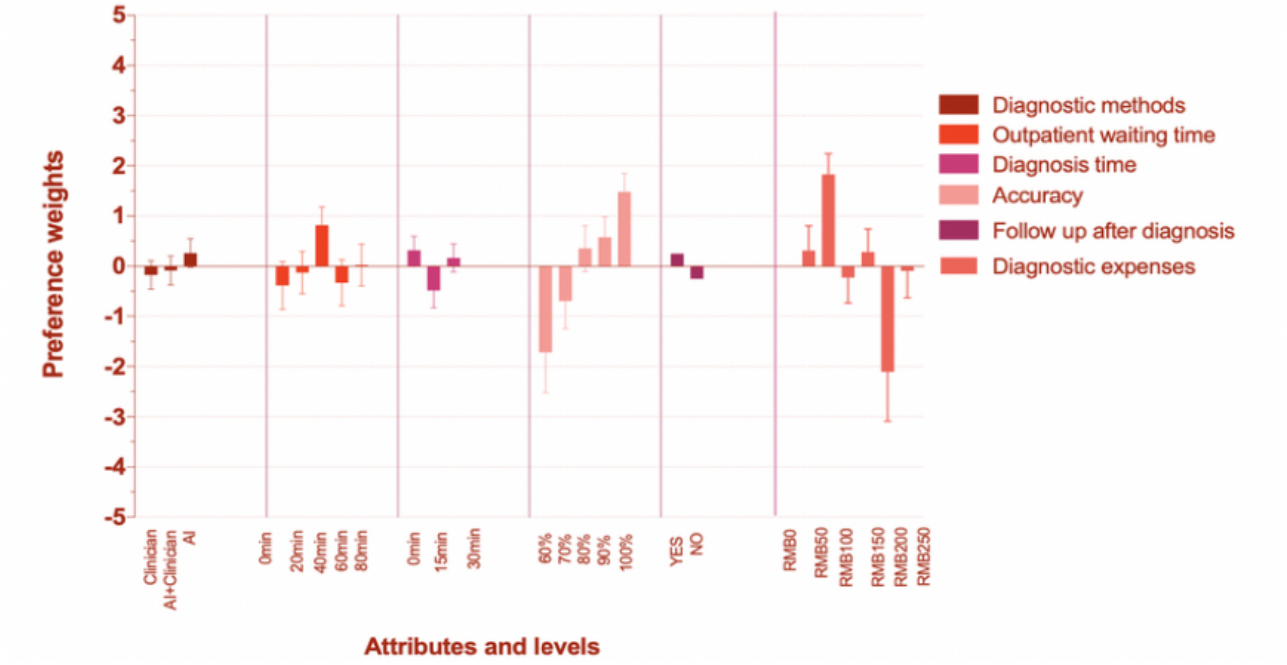
[a]OR: odds ratio.

[b]N/A: not applicable.

[c]A currency exchange rate of ¥1=US $0.15 is applicable.

For class 2, the attribute "accuracy," as well as "diagnostic expenses," was relatively important. Meanwhile, "diagnostic expenses" was (surprisingly) the most important for respondents among the attributes with a percentage weight of 39.09%, followed by "accuracy" with a percentage of 31.75% (Figure 1). The span of the percentage weights of these two attributes has been presented in Figure 3. The percentage weights for attribute "accuracy" were from −1.717 to 1.480, and −2.10 to 1.830 for "diagnostic expenses" (Figure 3). From *P* values, we find that almost all the levels are not statistically significant except the "100%" level of "accuracy" and level "¥50" of "diagnostic expenses." The odds ratio (Table 7) displays that the AI method is the best of three methods. The odds ratio of levels of "outpatient waiting time" are all greater than 1, and patients' preference weight still increases with the increasing of "accuracy" as in class 1. For attribute "diagnostic expenses," only the level "¥50" has an odds ratio greater than 1.

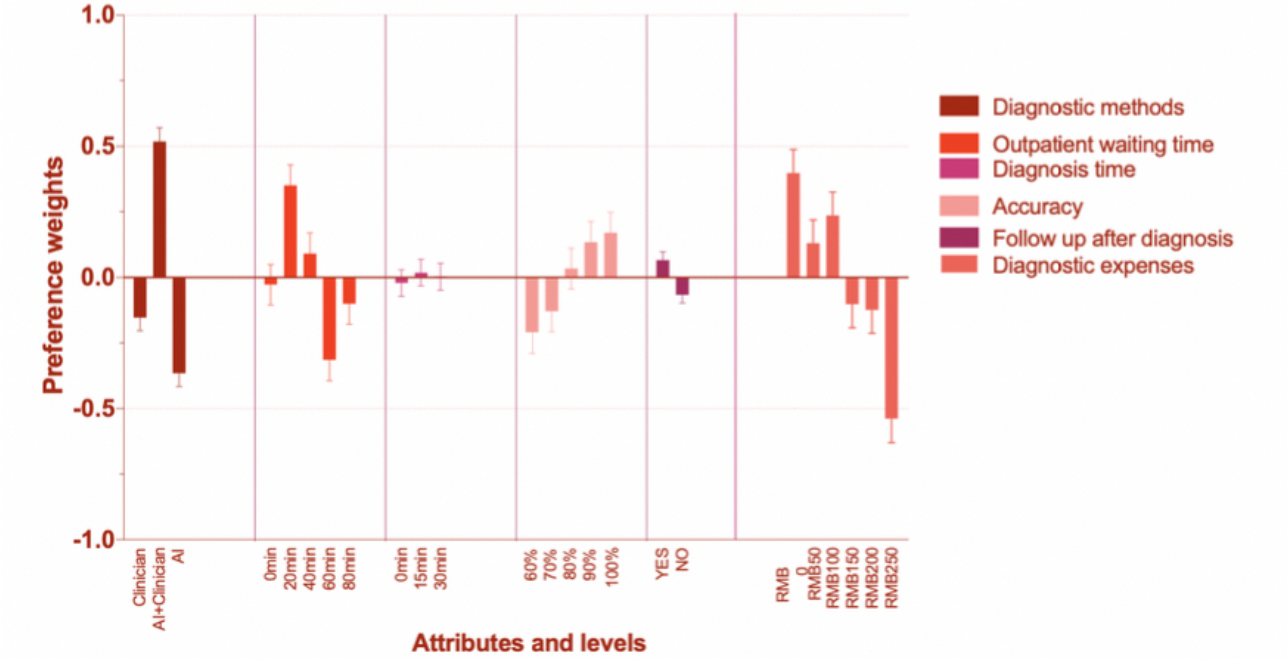
Figure 3. Latent class percentage weights in class 2. AI: artificial intelligence; RMB: yuan renminbi.



For class 3, attributes "diagnosis method" and "outpatient waiting time" were more statistically significant across all the attributes. For respondents in this class, the percentage importance of "diagnosis method" and "outpatient waiting time" were 29.09% and 30.83%, respectively (Figure 1). From Figure 4, we find that the span for diagnostic methods is from −0.364 to 0.518. Meanwhile, the span for diagnostic expenses is from −0.538 to 0.397. In addition, the odds ratio of level "AI + clinicians" of attribute "diagnostic method" was 1.985 (95% CI 1.769-2.167), which is greater than 1, meaning that respondents in class 3 also prefer durable diagnosis mode rather than single mode, similar to the condition of class 1. All of the odds ratios of levels of attributes "outpatient waiting time" and "accuracy" are greater than 1. Meanwhile, the odds ratio continues increasing with the increase of accuracy, the same condition of the previous two classes.

For these 3 latent classes, attribute "accuracy" is the most preferred factor in two classes (class 1 and class 2), while "diagnostic expenses" is the most preferred factor in class 3. In addition, the odds ratio of levels of attribute "accuracy" always increases with the increasing of accuracy rate, meaning that people's preference will always grow with the increase, and higher accuracy is always preferred.

Figure 4. Latent class percentage weights in class 3. AI: artificial intelligence; RMB: yuan renminbi.



Discussion

Principal Findings

Several different models were used in our study to research people's preference for various attributes, including a GMNL model and LCM; both models presented with their own merits and shortcomings.

For the GMNL analysis and the utility report, we found that "accuracy" was the most important thing for respondents and most of the levels of attributes "diagnostic methods" and "accuracy" were statistically significant. From the odds ratio, most of the respondents have a preference for the AI diagnosis plus clinicians' confirmation with outpatient waiting time of 20 minutes, diagnosis time of 15 minutes, and 100% accuracy. We would expect that the outpatient waiting time of 0 minutes and diagnosis time of 0 minutes should be preferred. However, the preferred choice was not 0 minutes. We assume that some people believe that a longer diagnosis time can give them more credibility and a feeling of safety. Furthermore, the odds ratio of 1.559 for "AI diagnosis + clinicians' confirmation" was evidently higher than only "AI diagnosis," which reflects that, for now, the vast majority of people still cannot totally trust the AI diagnosis due to the factors of AI uncertainty. AI diagnosis acting in concert with clinicians will significantly guarantee the accuracy. Another defect of AI diagnosis is that patients follow up after diagnosis, which means that patients who cannot receive follow-up have no alternatives but to head to clinicians as outpatients, which indirectly results in missing the best patient treatment time. That also leads to huge inconvenience to both patients and clinicians. Those who have an evident preference for only AI diagnosis may as a result of freshness and relatively higher diagnosis time of AI method, with sacrifice of accuracy in their mind. Thus, we can still consider the clinicians' diagnosis to remain irreplaceable, at least in the short term. The limitation of the GMNL model was also obvious; some respondents with preferences for other diagnostic methods and different diagnosis expenses cannot be reflected.

From the LCM, our study finds that respondents who are divided into 3 classes show different preferences and different patient profiles. There was slight heterogeneity compared to the GMNL model. Specifically, respondents in class 1 and class 2 still attach importance to "'accuracy"; however, respondents in class 3 mostly pay attention to "diagnostic expenses." Although medical insurance and accessibility are quite advanced today, some patients hardly get access to basic medical diagnosis or treatment, particularly in some underdeveloped or remote areas in China. Some old people would rely on their own self-healing function or immune system [33] rather than go to the hospital due to their outdated concepts of unaffordable diagnosis and treatment expenses. Therefore, the acceptability of AI diagnosis or even modern medical techniques among old people is palpably lower than among new generations in China due to the concepts of cost. In addition, we hypothesized that the numbers of hospitals in remote areas is undoubtedly lower than those in urban or advanced areas, which results in the people living in remote areas having to undertake the transportation time and cost. The long transit time and cost could also sometimes be fatal to these people since that would also force them to stay at home and miss the most appropriate diagnosis and treatment time. To sum up, the promotion and spread of AI diagnosis cannot ignore the need to set a suitable diagnosis price or give some discount and bonus according to the wealth status of patients.

Several previous studies [13] have found that most patients believe the outpatient waiting time plays a vital role in their decision-making behavior when faced with various choices of hospitals and clinicians. In addition, few studies [14] have attached importance to the quality and quantity of follow-up after clinical diagnosis. However, all of these studies ignored the effect of diagnosis accuracy and diagnosis expense in patients' trade-offs. Especially in the era of artificial intelligence, there are few studies addressing the accuracy of AI versus human clinicians rather than the outpatient waiting time or other factors. With diagnostic accuracy, clinically, the AI system can be programmed to probe and mark out some cancer indications such as prostate cancer and is more accurate than experts [34]. The clinical application could reduce pathology workload in this epidemic and in future clinical work. An AI system with expert-level grading performance might contribute a second opinion, aid in standardizing grading, and provide expert advice in areas with poor sanitation. As cloud computing capabilities come closer to life, overcomes limited memory, and central processing unit power [35]. We believe ever more and more people will trust the rapid diagnostic capabilities of AI. There is a lot of SARS-CoV-2–related scientific research undertaken to use AI to combat this pandemic by deploying new methods in the development of vaccines and drugs, as well as for public awareness [36]. Relying on the advantages of AI for medical auxiliary diagnosis, image analysis, remote consultation, etc during the outbreak of COVID-19, many AI devices have been used in first-line medicine. Moreover, AI is reducing cross infection. It has played an important role in therapeutic innovation. Health QR Code, which is a fusion of AI and big data, is a mobile phone app for everyone and uses red, yellow, and green colors to provide simple and effective intelligent services for personnel communication and economic and trade exchanges in the China in the "postepidemic era."

AI will continue to play an increasingly important role in controlling the public health crisis to save lives and economic recovery. AI contact tracing based on mobile communication technology becomes more mature [37]. The AI systems combining computed tomography and clinical symptoms can help to quickly diagnose SARS-CoV-2 patients [10]. A surrogate rapid diagnosis technique that is based on a deep learning neural network can be applied for discovering SARS-CoV-2 by analyzing the visual chest radiography imaging of patients [38].

Limitations

Limitations for Our DCE

Theoretically, the larger the sample size is, the smaller the changes we will find in the DCE; however, due to various reasons related to the pandemic, we applied convenience sampling in the data collection procedure. Thus, our sample size is relatively small and underrepresentative. Furthermore, the other significant shortcoming of our DCE data is that our

statistics do not stand for the point of view of all the people in China or other people worldwide due to the limited transmission of our questionnaire. We did not set the diseases as acute or chronic, which is an exogenous factor that will affect people's choice of waiting time.

Limitations for AI Diagnosis Propaganda

With AI diagnosis, there exists a requirement for rapid delivery and logistics of medicine. Furthermore, the conceptual propaganda for AI diagnosis and treatment are still not in place, particularly in some rural areas and among some old people with relatively traditional medical concepts. Confidence and trust in the AI diagnostic method still have a long way to go.

Conclusion

Segments of patients' preferences for these diagnosis options seem to be homogenous and convergent. All the attributes hypothesized and attributes' levels are evidently not ignorable during the implementation and widespread use of AI diagnosis techniques. People's preference for "accuracy" was obvious across different classes. Although "online treatment" has become more common today, accuracy has been sacrificed in exchange for so-called convenience, which is totally unwise. In addition, AI diagnosis technique developers as well as technology sellers, including hospitals, should take diagnosis expense into consideration and make pricing rules more flexible in the light of areas' economic development and individual patients' wealth status.

AI will definitely have a potential market and bright future, especially in the ongoing COVID-19 pandemic, since the AI diagnostic technology can ease the requirements of professional clinicians worldwide, particularly in rural areas.

Acknowledgments

FH contributed to the Introduction and Conclusion sections. BA and CJPZ have proposed revisions to the manuscript. WKM conceived the original idea, designed the study, and supervised the project. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

Conflicts of Interest

None declared.

References

1. Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf 2014 Sep;23(9):727-731 [FREE Full text] [doi: 10.1136/bmjqs-2013-002627] [Medline: 24742777]
2. World Health Organization. Health workforce requirements for universal health coverage and Sustainable Development Goals: Human Resources for Health Observer, Issue No. 17. 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/250330/9789241511407-eng.pdf [accessed 2021-01-20]
3. The World Bank. Physicians (per 1,000 people). The World Bank DataBank.: The World Bank URL: https://data.worldbank.org/indicator/SH.MED.PHYS.ZS
4. Lewis SM, Cullinane FM, Carlin JB, Halliday JL. Women's and health professionals' preferences for prenatal testing for Down syndrome in Australia. Aust N Z J Obstet Gynaecol 2006 Jun;46(3):205-211. [doi: 10.1111/j.1479-828X.2006.00567.x] [Medline: 16704474]
5. Bishop AJ, Marteau TM, Armstrong D, Chitty LS, Longworth L, Buxton MJ, et al. Women and health care professionals' preferences for Down's Syndrome screening tests: a conjoint analysis study. BJOG 2004 Aug;111(8):775-779. [doi: 10.1111/j.1471-0528.2004.00197.x] [Medline: 15270923]
6. Greene J, Hibbard JH, Sacks R, Overton V, Parrotta CD. When patient activation levels change, health outcomes and costs change, too. Health Aff (Millwood) 2015 Mar;34(3):431-437. [doi: 10.1377/hlthaff.2014.0452] [Medline: 25732493]
7. Viswanathan M, Kraschnewski JL, Nishikawa B, Morgan LC, Honeycutt AA, Thieda P, et al. Outcomes and costs of community health worker interventions: a systematic review. Med Care 2010 Sep;48(9):792-808. [doi: 10.1097/MLR.0b013e3181e35b51] [Medline: 20706166]
8. Kennedy ADM, Sculpher MJ, Coulter A, Dwyer N, Rees M, Abrams KR, et al. Effects of decision aids for menorrhagia on treatment choices, health outcomes, and costs: a randomized controlled trial. JAMA 2002 Dec 04;288(21):2701-2708. [doi: 10.1001/jama.288.21.2701] [Medline: 12460093]
9. Ministry of Industry and Information Technology of China. Increase the application of artificial intelligence in precise prevention and control of epidemics. 2020. URL: http://www.gov.cn/xinwen/2020-02/19/content_5480997.htm [accessed 2021-01-20]
10. Mei X, Lee H, Diao K, Huang M, Lin B, Liu C, et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med 2020 Aug;26(8):1224-1228 [FREE Full text] [doi: 10.1038/s41591-020-0931-3] [Medline: 32427924]
11. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. Radiology 2020 Aug;296(2):E65-E71 [FREE Full text] [doi: 10.1148/radiol.2020200905] [Medline: 32191588]

12. Mollalo A, Rivera KM, Vahedi B. Artificial Neural Network Modeling of Novel Coronavirus (COVID-19) Incidence Rates across the Continental United States. Int J Environ Res Public Health 2020 Jun 12;17(12):4204 [FREE Full text] [doi: 10.3390/ijerph17124204] [Medline: 32545581]
13. Fetter R, Thompson JD. Patients' waiting time and doctors' idle time in the outpatient setting. Health Serv Res 1966;1(1):66-90 [FREE Full text] [Medline: 5915335]
14. De Stefano V, Leone G, Mastrangelo S, Tripodi A, Rodeghiero F, Castaman G, et al. Clinical manifestations and management of inherited thrombophilia: retrospective analysis and follow-up after diagnosis of 238 patients with congenital deficiency of antithrombin III, protein C, protein S. Thromb Haemost 1994 Sep;72(3):352-358. [Medline: 7855783]
15. McFadden D, Train K. Mixed MNL models for discrete response. J Appl Econ 2000 Sep;15(5):447-470. [doi: 10.1002/1099-1255(200009/10)15:5<447::aid-jae570>3.0.co;2-1]
16. Fiebig DG, Keane MP, Louviere J, Wasi N. The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. Marketing Science 2010 May;29(3):393-421. [doi: 10.1287/mksc.1090.0508]
17. David A H, William H G. The Mixed Logit model: The state of practice. Transportation 2003 May;30:133-176.
18. Pancras J, Dey DK. A comparison of generalized multinomial logit and latent class approaches to studying consumer heterogeneity with some extensions of the generalized multinomial logit model. Appl Stochastic Models Bus Ind 2011 Dec 21;27(6):567-578. [doi: 10.1002/asmb.944]
19. Garrett E, Zeger SL. Latent class model diagnosis. Biometrics 2000 Dec;56(4):1055-1067. [doi: 10.1111/j.0006-341x.2000.01055.x] [Medline: 11129461]
20. Roeder K, Lynch KG, Nagin DS. Modeling Uncertainty in Latent Class Membership: A Case Study in Criminology. Journal of the American Statistical Association 1999 Sep;94(447):766-776. [doi: 10.1080/01621459.1999.10474179]
21. Ben-Akiva M, Lerman SR. Discrete Choice Analysis: Theory and Application to Travel Demand. Cambridge, MA: MIT Press; 1985.
22. Berhane A, Enquselassie F. Patients' preferences for attributes related to health care services at hospitals in Amhara Region, northern Ethiopia: a discrete choice experiment. Patient Prefer Adherence 2015;9:1293-1301 [FREE Full text] [doi: 10.2147/PPA.S87928] [Medline: 26396501]
23. Alizadeh A, Eftekhaari TE, Mousavi SH. Patient preferences for hospital quality in Bandar Abbas using a discrete choice experiment: 2010-2011. Life Science Journal 2012;9(4):1882-1886.
24. Ryan M. Using conjoint analysis to take account of patient preferences and go beyond health outcomes: an application to in vitro fertilisation. Soc Sci Med 1999 Feb;48(4):535-546. [doi: 10.1016/s0277-9536(98)00374-8] [Medline: 10075178]
25. Cheung GW, Rensvold RB. Testing Factorial Invariance across Groups: A Reconceptualization and Proposed New Method. Journal of Management 2016 Jun 30;25(1):1-27. [doi: 10.1177/014920639902500101]
26. Bailey RA. Balance, Orthogonality and Efficiency Factors in Factorial Design. Journal of the Royal Statistical Society: Series B (Methodological) 2018 Dec 05;47(3):453-458. [doi: 10.1111/j.2517-6161.1985.tb01374.x]
27. Fiebig DG, Hall J. 6 Discrete choice experiments in the analysis of health policy. In: Quantitative Tools for Microeconomic Policy Analysis. 2005 Presented at: Quantitative Tools for Microeconomic Policy Analysis; November 17-18, 2004; Canberra, Australia p. 119-136.
28. Greene WH, Hensher DA. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B: Methodological 2003 Sep;37(8):681-698. [doi: 10.1016/s0191-2615(02)00046-2]
29. Sakamoto Y, Ishiguro M, Kitagawa G. In: Reidel D, editor. Akaike Information Criterion Statistics. Dordrecht, the Netherlands: D. Reidel; 1986.
30. deLeeuw J. Introduction to Akaike (1973) information theory and an extension of the maximum likelihood principle. In: Breakthroughs in statistics. New York, NY: Springer; 1992:599-609.
31. Weakliem DL. A Critique of the Bayesian Information Criterion for Model Selection. Sociological Methods & Research 2016 Jun 30;27(3):359-397. [doi: 10.1177/0049124199027003002]
32. Sclove SL. Application of model-selection criteria to some problems in multivariate analysis. Psychometrika 1987 Sep;52(3):333-343. [doi: 10.1007/bf02294360]
33. Rowen D, Stevens K, Labeit A, Elliott J, Mulhern B, Carlton J, et al. Using a Discrete-Choice Experiment Involving Cost to Value a Classification System Measuring the Quality-of-Life Impact of Self-Management for Diabetes. Value Health 2018 Jan;21(1):69-77 [FREE Full text] [doi: 10.1016/j.jval.2017.06.016] [Medline: 29304943]
34. Ström P, Kartasalo K, Olsson H, Solorzano L, Delahunt B, Berney DM, et al. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. Lancet Oncol 2020 Feb;21(2):222-232. [doi: 10.1016/S1470-2045(19)30738-7] [Medline: 31926806]
35. Tawalbeh LA, Mehmood R, Benkhlifa E, Song H. Mobile Cloud Computing Model and Big Data Analysis for Healthcare Applications. IEEE Access 2016;4:6171-6180. [doi: 10.1109/access.2016.2613278]
36. Ahuja AS, Reddy VP, Marques O. Artificial intelligence and COVID-19: A multidisciplinary approach. Integr Med Res 2020 Sep;9(3):100434 [FREE Full text] [doi: 10.1016/j.imr.2020.100434] [Medline: 32632356]
37. Kricka L, Polevikov S, Park J, Fortina P, Bernardini S, Satchkov D, et al. Artificial Intelligence-Powered Search Tools and Resources in the Fight Against COVID-19. EJIFCC 2020 Jun;31(2):106-116 [FREE Full text] [Medline: 32549878]

38. Panwar H, Gupta P, Siddiqui MK, Morales-Menendez R, Singh V. Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. Chaos Solitons Fractals 2020 Sep;138:109944 [FREE Full text] [doi: 10.1016/j.chaos.2020.109944] [Medline: 32536759]

Abbreviations

ABIC: sample size–adjusted BIC
AI: artificial intelligence
AIC: Akaike information criterion
BIC: Bayesian information criterion
CBC: choice-based conjoint
DCE: discrete choice experiment
GMNL: generalized multinomial logit
LCM: latent class model
MXL: mixed logit
OR: odds ratio
S-MNL: scaled-multinomial logit
WHO: World Health Organization

Edited by G Eysenbach, R Kukafka; submitted 24.07.20; peer-reviewed by A Al-Hasan; comments to author 03.09.20; revised version received 15.09.20; accepted 20.01.21; published 23.02.21.

Please cite as:
Liu T, Tsang W, Huang F, Lau OY, Chen Y, Sheng J, Guo Y, Akinwunmi B, Zhang CJP, Ming WK
Patients' Preferences for Artificial Intelligence Applications Versus Clinicians in Disease Diagnosis During the SARS-CoV-2 Pandemic in China: Discrete Choice Experiment
J Med Internet Res 2021;23(2):e22841
URL: https://www.jmir.org/2021/2/e22841
doi:10.2196/22841
PMID:33493130

Original Paper

Public Opinions and Concerns Regarding the Canadian Prime Minister's Daily COVID-19 Briefing: Longitudinal Study of YouTube Comments Using Machine Learning Techniques

Chengda Zheng[1*], BSc; Jia Xue[1,2*], PhD; Yumin Sun[1], BSc; Tingshao Zhu[3,4], PhD

[1]Faculty of Information, University of Toronto, Toronto, ON, Canada

[2]Factor Inwentash Faculty of Social Work, University of Toronto, Toronto, ON, Canada

[3]Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

[4]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Tingshao Zhu, PhD
Key Laboratory of Behavioral Science
Institute of Psychology
Chinese Academy of Sciences
16 Lincui Road, Chaoyang district
Beijing, 100101
China
Phone: 86 1064871661
Email: tszhu@psych.ac.cn

Abstract

Background: During the COVID-19 pandemic in Canada, Prime Minister Justin Trudeau provided updates on the novel coronavirus and the government's responses to the pandemic in his daily briefings from March 13 to May 22, 2020, delivered on the official Canadian Broadcasting Corporation (CBC) YouTube channel.

Objective: The aim of this study was to examine comments on Canadian Prime Minister Trudeau's COVID-19 daily briefings by YouTube users and track these comments to extract the changing dynamics of the opinions and concerns of the public over time.

Methods: We used machine learning techniques to longitudinally analyze a total of 46,732 English YouTube comments that were retrieved from 57 videos of Prime Minister Trudeau's COVID-19 daily briefings from March 13 to May 22, 2020. A natural language processing model, latent Dirichlet allocation, was used to choose salient topics among the sampled comments for each of the 57 videos. Thematic analysis was used to classify and summarize these salient topics into different prominent themes.

Results: We found 11 prominent themes, including strict border measures, public responses to Prime Minister Trudeau's policies, essential work and frontline workers, individuals' financial challenges, rental and mortgage subsidies, quarantine, government financial aid for enterprises and individuals, personal protective equipment, Canada and China's relationship, vaccines, and reopening.

Conclusions: This study is the first to longitudinally investigate public discourse and concerns related to Prime Minister Trudeau's daily COVID-19 briefings in Canada. This study contributes to establishing a real-time feedback loop between the public and public health officials on social media. Hearing and reacting to real concerns from the public can enhance trust between the government and the public to prepare for future health emergencies.

(J Med Internet Res 2021;23(2):e23957) doi:10.2196/23957

KEYWORDS

Canada; PM Trudeau; YouTube; machine learning; big data; infodemiology; infodemic; public concerns; communication; concern; social media; video

Introduction

The World Health Organization declared COVID-19 to be a global public health emergency on March 11, 2020 [1]. Canada confirmed its first COVID-19 case on January 25, 2020, and a total of 189,387 cases of COVID-19 in Canada were reported as of October 14, 2020 [2]. On March 12, 2020, Quebec was the first province to declare a state of emergency [3], and the government established preventive measures such as transportation restrictions, quarantine rules, and social distancing requirements [4,5]. In Canada, Prime Minister Justin Trudeau started providing a daily briefing on the government's updated policies and actions for stopping the spread of COVID-19 on March 13, 2020. For example, Prime Minister Trudeau announced an interest-free moratorium on student loan payments and closure of the United States–Canada border on March 18, 2020; announced the Canada Emergency Response Benefit (CERB) on March 25; and announced the Canada Emergency Student Benefit (CESB) through independent public entities (eg, lowering of the interest rate by the Bank of Canada) or federal government programs (eg, CERB and CESB) on April 22 [6].

Since the COVID-19 outbreak, social media has become the most accessible source for obtaining news and health information and for exchanging opinions. Given their self-isolation, people extensively post opinions, express emotions, and exchange ideas about COVID-19–related policies on the web [7]. In this study, we analyzed the YouTube comments under Prime Minister Trudeau's daily briefing videos. YouTube is the first web-based video sharing platform. With more than 2 billion users and 1 billion hours of videos watched every day, YouTube has become one of the largest video sharing platforms worldwide, and it has played an essential role in public communication during the COVID-19 epidemic [8]. Comments on YouTube have provided rich data for public discourse and sentiment research. Existing studies examine the impact of YouTube videos and comments on epidemic diseases (eg, Ebola virus, Zika virus) and show that YouTube videos and comments are essential channels for disseminating health information [9-11]. Basch et al [9] coded the source and contents of the 100 most widely viewed videos about the Ebola virus on YouTube. They found that these videos have been viewed 73 million times and included contents about the death toll of the virus and how it was transmitted. Pathak et al [11] evaluated 198 YouTube videos about the Eloba outbreak in 2014. They found that most of the videos were useful, and they demonstrated that YouTube is a useful health information source during global health emergencies. In contrast, Bora et al [10] concluded that videos could potentially spread misinformation during global public health emergencies, and trustworthy videos from health organizations or universities were scarce. Khatri and colleagues [12] analyzed content about COVID-19 on YouTube by reviewing 72 videos in English and 42 videos in Mandarin; they demonstrated that YouTube is an important platform for health information dissemination.

This study has two goals. The first is to examine whether YouTube comments are a useful source of public opinions and priorities on COVID-19; the second is to identify public responses and discourses related to Prime Minister Trudeau's COVID-19 policies in Canada over time. According to Bora et al [10], we purposively selected trustworthy videos published by the official YouTube channel of the Canadian Broadcasting Corporation (CBC). More specifically, this study examines and tracks public discourse under Prime Minister Trudeau's daily COVID-19 briefing videos posted on YouTube (N=57) from March 13 to May 25, 2020, in Canada. This paper contributes to understanding the undermeasured public responses to Prime Minister Trudeau's COVID-19–related policies in Canada. Real-time and longitudinal analysis of public responses and concerns can help public health authorities recognize Canada's public priorities.

Methods

We followed the text mining pipeline, including data preparation and data analysis [7]. Data preparation included data sampling, collection, and preprocessing. Data analysis included unsupervised machine learning and thematic analysis. The unit of analysis was each unique comment posted under Prime Minister Trudeau's daily briefing videos from March 13 to May 25, 2020.

Sampling

We used a purposive sampling approach. Our sampling frame was Prime Minister Trudeau's COVID-19 daily briefing videos, which were published on the official CBC channel on YouTube [13] starting on March 13, 2020. Our sampling frame included 58 daily briefing videos posted from March 13 to May 22, 2020 (57 videos included comments; 1 did not receive any comments).

We present the characteristics of these videos in Multimedia Appendix 1, including publication dates; titles and descriptions of the videos; and numbers of views, likes, dislikes, and associated comments that each video received up to May 25, 2020. For example, on March 22, 2020, the daily briefing with the title "PM Trudeau announced that the $82-billion financial package was just the first step, and more financial funds will come" received the highest number of comments (n=2413) among the daily briefing videos published in March. The video titled "The term 'New Normal' to describe daily routine during the COVID-19 pandemic and the numbers would take months of continued, determined effort" had the highest number of comments (n=2087) in April. The video titled "Trudeau urged world leaders to pull together for the COVID-19 vaccine" received the highest number of comments (n=2653) in May.

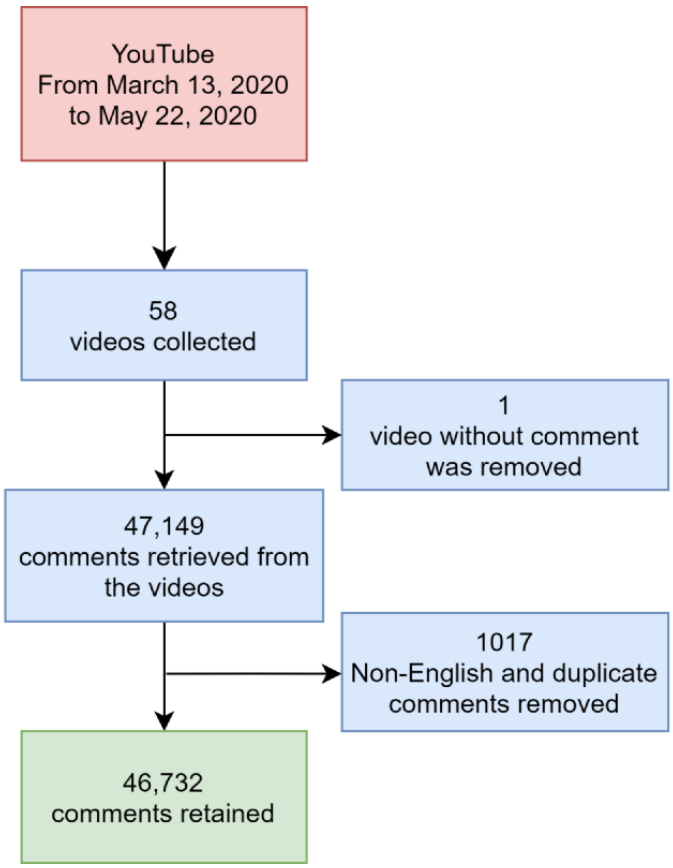
Data Collection

The data set of the study consisted of all the comments posted by YouTube users under each of the 57 videos. We used YouTube's open application programming interface to collect all the comments under each of Prime Minister Trudeau's daily briefing videos on May 25, 2020. We coded the information retrieval script in Python, version 3 (Python Software Foundation). We retrieved a total of 47,149 comments from these 57 videos. After removing 1017 non-English and duplicate comments, 46,732 comments remained as our final data set, as shown in Figure 1. The following information associated with each of the 46,732 comments was collected: (1) a unique ID

related to each comment; (2) the full comment text; (3) metadata associated with the comment, such as the reply count (number of replies that comment received), like count (number of the likes that comment received), published date and time of the comment, and deidentified username. The collected 46,732 English comments included both the first level comment and its associated "replyTo" (comment posted in reply to another comment) from 17,211 YouTube users.

We treated the set of comments published under each video as a separate document for the topic modeling (eg, documents 1-57). Thus, we obtained a total of 57 documents for topic modeling.

Figure 1. Data preprocessing chart.



Preprocessing of the Raw Data

We preprocessed the raw data to ensure data analysis quality. Preprocessing is an important task for analyzing YouTube comments, as it cleans the data by diminishing its complexity to provide better results [14]. Our data preprocessing included the steps below.

1. We removed handles (@user) and their content, as they did not contribute to the analysis.
2. We removed all non-English characters (non–American Standard Code for Information Interchange [ASCII] characters) because this study focused on English messages.
3. We removed words that did not make sense, such as "#newYork sdaaawd asjdasd @!!!!" We used the Nostril evaluator [15] to detect nonsense comments and pronunciations. Nostril incorporates a large table of n-gram frequencies to support its probabilistic assessment of text strings. For the comments labeled as "nonsense" by Nostril, we also checked their pronunciations (ie, vowels and phones). The detected nonsense comments were removed from the analysis.
4. We removed stopwords to ensure better results [16]. Stopwords (eg, the, a, an, in) are frequently occurring words that do not carry any information or orientation. We used the Natural Language Toolkit suite of Python libraries to filter out the stopwords to save space in our database and shorten the processing time.
5. We applied a stemmer. A stemmer is a rule-based process of stripping the suffixes (eg, "ing," "ly," "es," "s") from words. A large number of conjugates were removed.

Data Analysis

Unsupervised Machine Learning

We used latent Dirichlet allocation (LDA) [17] to generate prominent topics. LDA is a generative probabilistic model based on the hierarchical Bayesian that mines the underlying set of topics on a corpus of text. LDA has been applied to COVID-19–relevant topics on social media. For example, Xue et al [7] examined Twitter posts related to the COVID-19 pandemic. Obadimu and colleagues [18] applied LDA to recognize the toxicity of comments on YouTube. In this study, we treated the collected YouTube comments as a document and applied LDA topic modeling with the gensim Python library. For each identified topic, we used WordNetLemmatizer to extract popular unigrams and bigrams. The pyLDV library was used to visualize the findings.

Qualitative Analysis

The thematic analysis enabled us to interpret the patterns, topics, and themes from the unsupervised machine learning processes [19,20]. Two authors independently assigned topics based on the bigrams and representative comments related to each of the videos. Then, they reviewed all the identified topics and their representative unigrams, bigrams, and comment examples from the 57 videos and assigned themes across these topics. The research team examined the initial coded themes and considered whether they reflected the identified topics. The team discussed the themes and provided names to ensure that they reflected the identified salient topics.

Results

Descriptive Results of the Videos

Our sampling frame included 57 of Prime Minister Trudeau's daily briefing videos published on the CBC YouTube channel between March 13 and May 22, 2020. We describe the statistics of the videos in Table 1, including the numbers of views, likes, dislikes, and comments of the 57 videos. For example, on average, these videos were viewed 364,487 times and received an average of 830 comments and 1826 likes.

Table 1. Descriptive statistics of the collected videos (N=57).

Characteristic	Range	Mean (SD)
Number of views	74,924-839,921	364,487 (205,984)
Likes	492-4405	1826 (1059)
Dislikes	290-2750	676 (344)
Comments	197-2653	830 (544)

Figure 2 shows the view counts of Prime Minister Trudeau's daily briefings over time. The view counts peaked on March 20 and gradually decreased after April, when Prime Minister Trudeau declared that the government would close the border to stop the spread of COVID-19 [21].

Figure 2. The number of views for the sampled daily briefing videos over time (n=56).

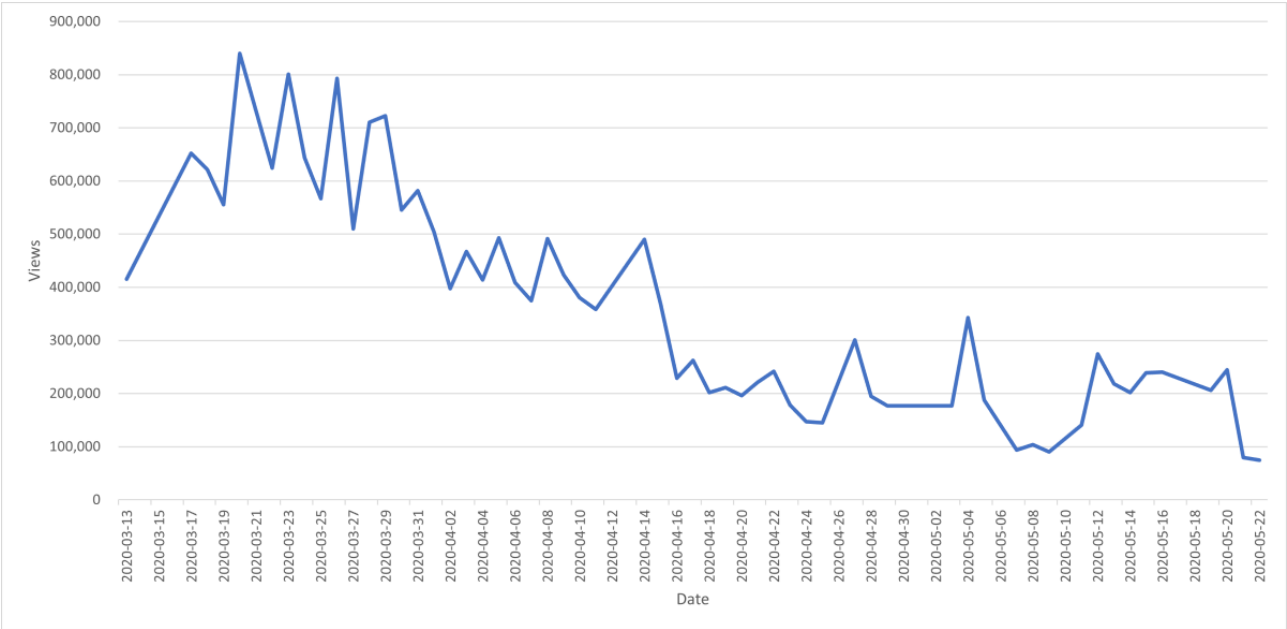
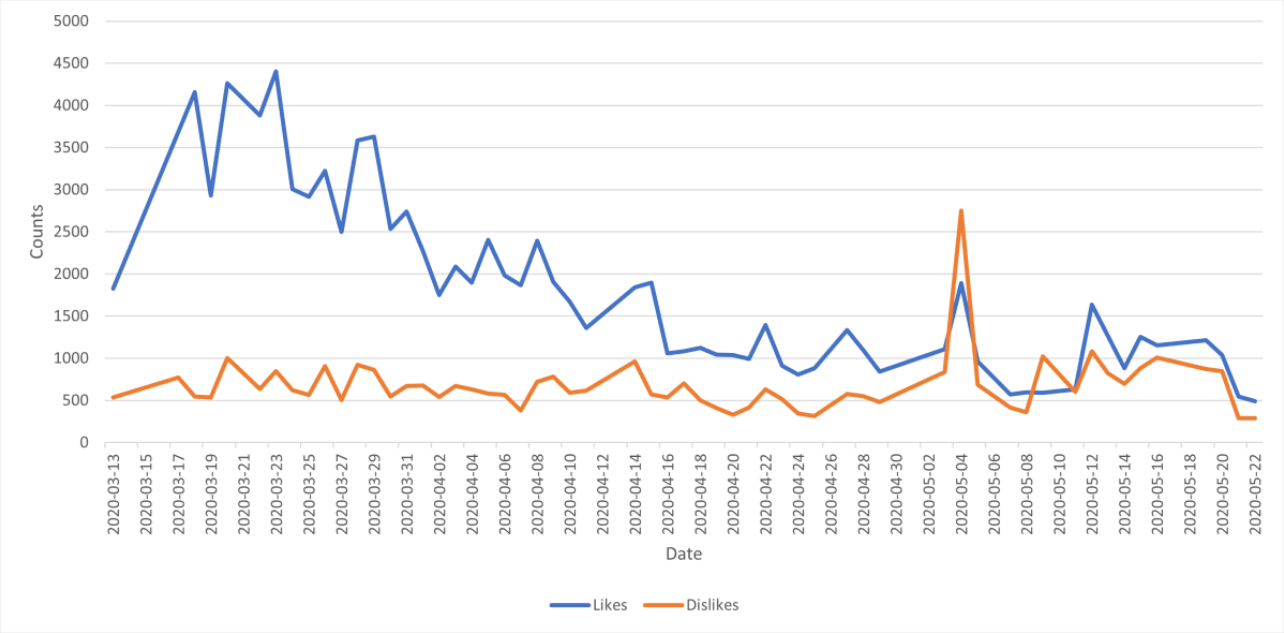


Figure 3 shows that the numbers of likes were higher than the numbers of dislikes on most of the days, except for May 4 and May 9, 2020. The daily briefing on May 4 urged all world leaders to work together and fight for a COVID-19 vaccine. Additionally, this video received the highest number of comments. Prime Minister Trudeau stated that Canadians would receive more financial benefits on May 9, 2020; this video received more dislikes than likes.

Figure 3. Like counts and dislike counts for the sampled daily briefing videos over time.



Prominent Topics in the Comments

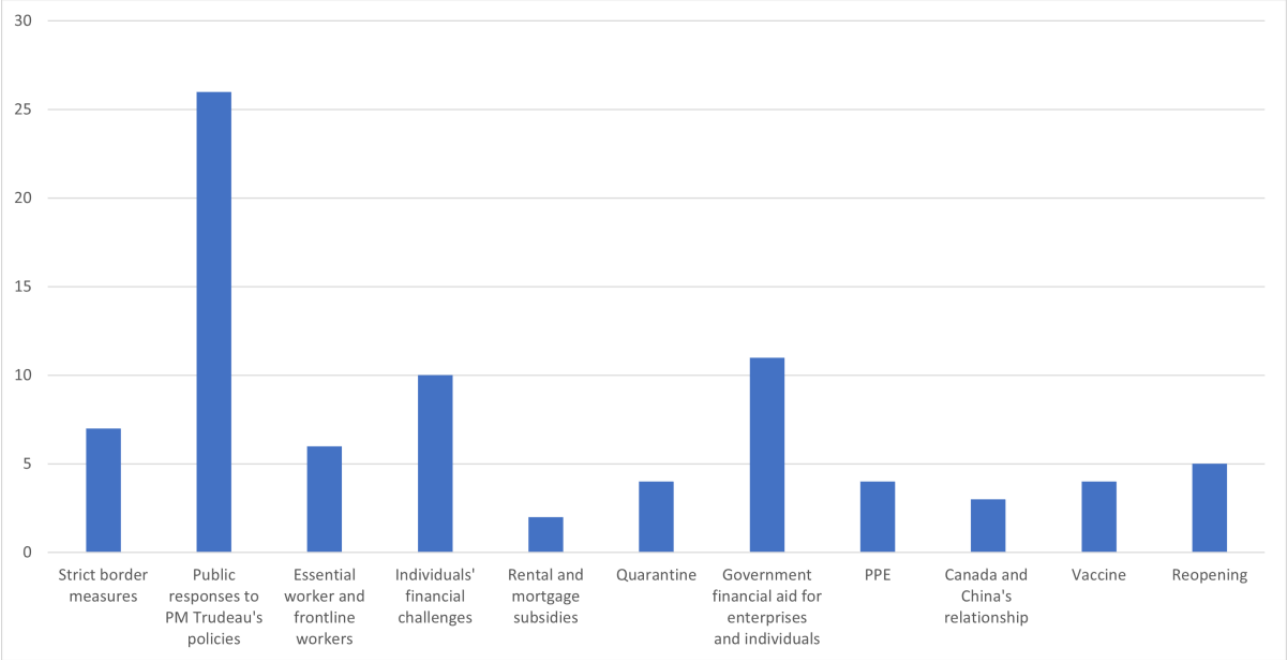
We retrieved and downloaded the associated 46,732 comments for each of the 57 videos as our final data set. We identified 1 to 3 salient topics for each of the documents (#1 to #57) associated with each daily briefing video, resulting in 168 topics. We present the topic modelling results for each of the 57 documents in Multimedia Appendix 2, including the publication dates of the videos, prominent topics in the documents, popular bigrams in the topics, and representative comments on each topic. For example, document #1 included all posted comments under Prime Minister Trudeau's daily briefing on March 13, 2020. We identified two prominent topics in document 1: border measures and comment on the government. Popular bigrams were "close the," "dam border," "international travel," and "prime minister." A representative YouTube comment is "CLOSE THE BORDER!!!! Wake up Trudeau… close the god dam border Now. Stop all flights from Italy and Europe NOW."

Prominent Themes

Multimedia Appendix 3 shows the salient themes with their descriptions, topics (bigrams), and representative comments. We identified 11 themes from the 168 most prominent topics, including strict border measures, public response to Prime Minister Trudeau's policies, essential work and frontline workers, individuals' financial challenges, rental and mortgage subsidies, quarantine, government financial aid for enterprises and individuals, personal protective equipment, Canada and China's relationship, vaccines, and reopening. For example, public response to Prime Minister Trudeau's policies was a prominent theme, exemplified by the topic (uni)bigrams of government, Prime Minister, Justin Trudeau, Canada, officials, liberal, government and right wing. An example comment is "Respect and listen to the Prime Minister, he is trying his best with the Government to keep everyone in our Country safe from Covid19."

Figure 4 shows the number of occurrences of the prominent themes between March 13 and May 22, 2020. For example, the theme of public responses to Prime Minister Trudeau's policies was the most popular, followed by government financial aid for enterprises and individuals and individuals' financial challenges.

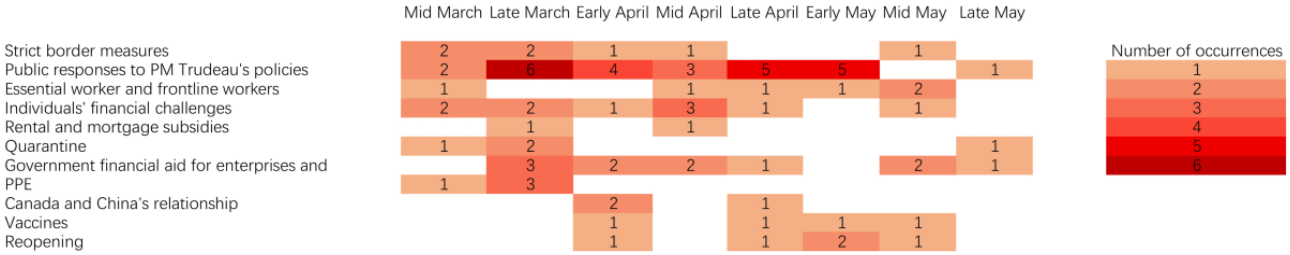
Figure 4. Occurrences of each theme from March 13 to May 22, 2020. PM: Prime Minister; PPE: personal protective equipment.



Themes Over Time

In addition to counting the number of occurrences, we also tracked the prominent themes over time, as shown in Figure 5. We divided each month into three periods: early month (day 1 to day 10), mid-month (day 11 to day 20), and late month (day 21 and after). Public concerns were ongoing for the themes of public responses to Prime Minister Trudeau's policies and individuals' financial challenges, as comments related to these themes appeared multiple times from March to May. For example, the theme of public responses to Prime Minister Trudeau's policies had been continuously discussed by the public since March (except for in mid-May). The theme of strict border measures mainly occurred from mid-March to mid-April. However, it was discussed again in mid-May. The themes of vaccines and reopening were mainly discussed from late April.

Figure 5. Prominence of each theme from March 13 to May 22, 2020. Early, day 1-10; mid-month, day 11-20; late, day 21 and after. PM: Prime Minister; PPE: personal protective equipment.



Popular Comments

We selected the 50 comments with the highest numbers of likes and the 50 comments with the highest numbers of replies. We present 19 comments that ranked in the top 50 most liked and replied in Table 2. The table shows the contents of these 19 comments along with their published dates, categorized themes, numbers of replies, and numbers of likes. For example, a single comment posted on March 28 received 488 likes and 104 replies from other YouTube users. This comment reflected the theme of strict border measures. Among the 19 selected popular comments, we found that five comments were about the theme of quarantine, and three comments reflected the theme of open government. The theme of quarantine was related to quarantine-related measures, as the public expressed a desire for more strict quarantine measures when they commented on and responded to Prime Minister Trudeau's daily briefings. The theme of open government was related to the public's request for a more open government and public interaction dynamics.

Table 2. Most popular comments, origins (daily briefing dates), assigned themes, and numbers of replies and likes.

Comment	Date of origin (2020)	Theme	Replies, n	Likes, n
"Why are we still not checking people at the airports?"	March 28	Strict border measures	104	488
"They should be pausing all mortgages and rent payments, the CRA can't give you money by the 1st"	March 22	Rental and mortgage subsidies	99	525
"I really hope when this is defeated that our society doesn't go more Orwellian"	May 04	Good wishes	88	230
"MR TRUDEAU: All you need to do is issue a moratorium stating that all Canadians with a rental lease or a mortgage is immediately suspended for 3 to 6 months. After which their existing monthly payments will restart and continue as usual. This will put at least 30% more money in every person's account fairly. These funds can be used to pay cred it cards, car loans, food etc. It is a brilliant idea already being implemented in the USA by a number of Banks and institutions. This is simple and would help all Canadians. Obviously no penalty would by involved and this would be a Government approved legal postponement of all rents and mortgages for the period stated."	March 24	Individuals' financial challenges	78	246
"Its time to come together and support each other and question everything."	May 04	Good wishes	68	802
"I don't care about staying home. I'm a homebody anyways. I care about being poor and ending up completely homeless after the whole Covid19 crisis ends. My landlord already wants me to move out by August, and no one knows if it will have died down by then or if I'll have a job by then so I can safely move."	March 29	Individuals' financial challenges	67	165
"Meanwhile, Canadian Tire, Lowes, Home Depot, Costco all the checkout attendants facing hundreds of customers per day and not one of them was wearing a mask and that was yesterday March 21st"	March 22	Quarantine	67	356
"Comments are allowed? Interesting"	March 26	Open government	66	335
"The scariest YouTube title ever"	April 09	Emotion	63	340
"I just want to know if these people are secretly getting haircuts. The rest of us will look like sheep dogs if this continues."	March 26	Emotion	54	178
"I wish you didnt do the voice over, he is already doing it bilingually. we dont need to hear it two times in English besides the translation is badly done."	April 14	Not related	54	615
"Anybody coming into Canada from other countries should have mandatory isolation for 14 days."	March 28	Quarantine	53	222
"I dont understand why people still went on cruises even tho they were specifically told not to a longgggggg time ago"	March 28	Emotion	51	269
"Lets have open questions from the public."	April 09	Open government	40	203
"Close the border dont exempt American citizens. Close it down."	March 17	Quarantine	38	304
"We should not rely on the US, like we are the 51st state!"	March 05	Emotion	34	180
"Dont have a social gettering of more than 10 people. Thats the guideline in South Korea. I think 50 is too much^^"	March 17	Quarantine	33	415
"Dr Tam waiting for 'evidence' of human to human transmission is disgraceful especially considering entire cities in China had already been locked down and the Three early German cases traced back to a Chinese business woman had already been reported. The precautionary principle should have been followed and Canadians deserve answers. We deserve better and Dr. Tam let us all down"	April 14	Quarantine	32	162
"Never in my life have I heard so many answers that dont answer a question. Ever. Lol"	April 14	Open government	32	317

Discussion

Principal Results

This study is the first to investigate and track public opinions and reactions to Canadian Prime Minister Trudeau's daily COVID-19 briefings. This study uses a machine learning approach and longitudinally examines public discourse about Prime Minister Trudeau's daily briefings over time. The results reveal several prominent themes over time, including strict border measures, public responses to Prime Minister Trudeau's policies, individuals' financial challenges, government financial aid for enterprises and individuals, and essential work and frontline workers. In addition, the themes of vaccines and reopening were highly discussed from late April. The study demonstrates that YouTube comments are a useful source of public opinions and priorities during global health emergencies.

Theme 1: Strict Border Measures

This theme refers to comments that discuss topics such as border entry policy, international travel, and border screening measures. The majority of the comments in this theme requested that the government implement stricter border measures to prevent transmission from carriers who have or do not have COVID-19 symptoms, exemplified by the comment "CLOSE THE BORDER... Stop all flights from Italy and Europe NOW." Public discourse on this topic started in mid-March and remained a hot topic until mid-April. The results correspond with a policy announced by the Canadian government on March 18, 2020, that implemented a ban on all foreign nationals entering Canada, including restrictions on all nonessential travel at the Canada–United States border. Discussions of border measures were not prominent in late April and early May but reappeared in mid-May, consistent with the border closure agreement between Canada and the United States, which was valid until May 21.

Our results suggest that the "border policy issue" has continuously been a concern for Canadians, who support strict border measures to prevent the spread of COVID-19 in Canada. Previous studies demonstrated that air transportation has the potential to spread influenza, such as A/H1N1 [22], and enhanced precautions should be taken for travelers to countries with increasing cases of respiratory diseases, such as Middle East respiratory syndrome [23]. Brown and colleagues [24] showed that airport exit and entry screening at international borders facilitated the rapid detection of illness and the implementation of appropriate public health control measures to prevent the spread of Ebola virus. Our results suggest that future research can provide empirical evidence for the public and policy makers regarding the effectiveness of border measures to prevent the spread of COVID-19, as this has been a primary public concern over time.

Theme 2: Public Responses to Prime Minister Trudeau's Policies

Public responses to Prime Minister Trudeau's policies refer to public opinion comments on the Canadian government, parties, or the Prime Minister himself. The majority of comments include opinions regarding the government's performance and policies during the COVID-19 pandemic. One representative comment is "Respect and listen the Prime Minister, he is trying his best with the Government to keep everyone in our Country safe from Covid19 with emergency funds for those that need it." Compared to other themes, this theme is more popular, and it continuously received public attention from mid-March to late May. The popularity of this theme is due to the positive interaction dynamic between the general public and the government of Canada, which encourages people to communicate with government organizations on social networks. This interaction can be traced back to 2011, when Canada acknowledged the need to use social media to interact with the public for the first time [25]. Social media users in Canada are primarily interested in communicating with the government and obtaining customized information [26]. One study [27] showed that social media engages the public to foster participatory dialogues and discussions of policy development and implementation. In addition, we found criticisms related to the government's COVID-19–relevant policies, such as:

> *Government Health Officials WHAT? They are months behind! Recommended NO MASKS only because they did not prepare, and they hadn't enough masks.*

Previous studies have examined mixed sentiments toward governments' policies in emergencies. During the 2009 H1N1 influenza pandemic, more than half the US population felt positive emotions toward the government's policy responses, while a certain number of people did not [28].

Theme 3: Individuals' Financial Challenges

Financial challenges refer to comments that include people's discussion, concerns, or policy proposals related to financial loss during the pandemic, such as "having no income for essential living costs due to lost jobs." One representative comment is:

> *I'm more worried about the financial issue than the health one right now. We are doing everything we can health-wise in this family, but my son is now out of work, and we know hubby will be soon.*

Our results show that the public has been frequently and continuously discussing finance-related topics since Prime Minister Trudeau's first daily briefing. The assessment of the impact of the COVID-19 pandemic on the labor market in Canada shows that the pandemic drove a 32% reduction in aggregate weekly hours worked from February to April 2020, along with a 15% reduction in employment [29]. As indicated by Statistics Canada, the average net savings for all Canadian households was CAD $852 (US $671) in 2018 [30], which suggests that people may be unable to afford essential living costs if they do not receive a basic income.

We found a positive correlation between the government's policies supporting individuals and public attention to financial challenges. CERB was launched on April 6, 2020; nearly 3.5 million Canadians applied for this benefit in the first week, and this number increased to 7.12 million on April 24, 2020 [31]. More than 7 million people received CAD $2000 (US $1575) from the Canadian government in April, which explains the temporary decrease in the public attention toward financial challenges in early May.

Theme 4: Government Financial Aid for Enterprises and Individuals

Discussion about enterprises and individual financial aid plans was prominent from March to May 2020. This theme includes discussions of government financial aid plans and programs supporting either enterprises or individuals, such as CERB, CESB, the Canada Emergency Wage Subsidy, and the Canada Emergency Business Account. This theme is different from theme 3, individuals' financial challenges, which focuses on individuals' concerns and discussions of their financial obstacles during the outbreak. In late March, the theme of government financial aid became prominent when the government administration firstly announced a number of economic stimulation programs, exemplified by this comment: "If your

business/source of income was interrupted due to COVID-19, whether you lost your business or not, you could still try to apply, as you would be in true need of help at this point." Our results confirm those of previous research on emergency individual financial aid during the pandemic and the importance of financial aid programs for businesses [32,33]. The longitudinal discussions on financial aid plans suggest that financial aid during the COVID-19 pandemic is critical for the public, and further research is suggested to assess the need for financial aid programs and evaluate their implementation.

Theme 5: Essential Work and Frontline Workers

Essential work and frontline workers was a prominent theme in the comments related to Prime Minister Trudeau's daily briefings, such as the welfare system in Canada, incentive pay for essential work, and protection of frontline workers. One representative comment stated, "The government should be incentivizing those who have to work during a pandemic, even a $500/month incentive." This was an active public discussion topic in mid-March; it lost public attention for 20 days, and then became prominent again in mid-April. Our results are consistent with an earlier study that shows that frontline workers should be supported, and public support of frontline workers has been related to a speech by the Prime Minster of India [34]. Our findings suggest that Prime Minister Trudeau's daily speech advocating for frontline workers also resulted in support and attention from the general public. For example, Prime Minister Trudeau mentioned essential work and frontline workers several times during his daily briefings, especially on April 29 and May 7. Our study suggests that political efforts in acknowledging essential workers has a substantial impact on the general public's awareness. It is worth examining whether public acknowledgment influences people's behaviors, such as strictly following social isolation and quarantine measures.

Theme 6: Vaccines

Public discussion of vaccines is another prominent theme that emerged in April. We classified all relevant topics in this theme, such as the development of vaccines, exemplified by this comment: "It could take years to come up with a vaccine, you've been doing a good job, Trudeau. Don't it up by making statements you can't back up." Previous studies analyzing Tweets related to COVID-19 also found that vaccines were a highly discussed topic [7]. Our findings provide further evidence that vaccines have been a significant public concern among Canadians during the COVID-19 outbreak. However, we also found misinformation regarding vaccines on social media, where some comments supported that vaccines are a type of conspiracy. Our findings suggest that the Canadian government should provide more resources to the public to dispel misinformation regarding vaccines. Proper public education is essential to raise public awareness and to mobilize and engage people in communities in Canada.

Limitations

There are several limitations to this study. First, this paper is not based on random sampling. We purposively collected comments that were posted by YouTube users under Prime Minister Trudeau's briefing videos. They only represent a portion of the public opinions regarding Prime Minister Trudeau's daily briefings during the COVID-19 pandemic, which causes biases. Second, our sample size is relatively small compared to the large amount of data on Twitter. Third, we treated secondary comments (ie, comment to comment) in the same manner as first-level comments for the study aims. Thus, this study has limitations in presenting the structural relationship between first-level and second-level comments. Finally, we retrieved all videos, comments, and associated data on May 25, 2020, three days after the last day of the Prime Minister's daily briefing. It is noted that the evaluations could be different due to the increased number of comments added over time.

Conclusion

This study is the first to analyze and track the public discourse of Canadian Prime Minister Justin Trudeau's daily briefing on COVID-19 on social media. Machine learning analysis was applied to 46,732 English YouTube comments from 57 videos from March 13 to May 22, 2020, and the results suggest that several prominent themes should be noted by public health agencies and policy makers, such as strict border measures, public responses to Prime Minister Trudeau's policies, individuals' financial challenges, and government financial aid for enterprises and individuals. Recommendations for future work include (1) further validating YouTube comments as a source of information on COVID-19 and (2) strategies that can be considered by government and public health agencies to strengthen the real-time feedback loop between those agencies and the public on YouTube. Hearing and reacting to real concerns from the public can enhance trust between the government and the public to prepare for future health emergencies.

Authors' Contributions

CZ and JX contributed equally as coauthors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Descriptive characteristics of the sampling frame (N=57 videos).
[DOCX File , 32 KB - jmir_v23i2e23957_app1.docx]

Multimedia Appendix 2
Modeling of prominent topics for YouTube comments on each daily briefing (video dates, topics, bigrams, and representative examples).
[DOCX File , 34 KB - jmir_v23i2e23957_app2.docx]

Multimedia Appendix 3
Salient themes, topic keywords, and representative comments.
[PNG File , 177 KB - jmir_v23i2e23957_app3.png]

References

1. Coronavirus disease (COVID-19) pandemic. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [accessed 2020-05-15]
2. Coronavirus disease (COVID-19). Government of Canada. URL: https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19.html [accessed 2021-02-11]
3. Financial assistance for businesses. Government of Québec. URL: https://www.quebec.ca/en/health/health-issues/a-z/2019-coronavirus/financial-assistance-covid19/ [accessed 2020-10-15]
4. Bronca T. COVID-19: A Canadian timeline. Canadian Healthcare Network. 2020 Apr 08. URL: https://www.canadianhealthcarenetwork.ca/covid-19-a-canadian-timeline [accessed 2020-10-15]
5. Migone AR. Trust, but customize: federalism's impact on the Canadian COVID-19 response. Policy Soc 2020 Jun 19;39(3):382-402. [doi: 10.1080/14494035.2020.1783788]
6. Lord P, Saad L. Outline of government programs related to the COVID-19 pandemic in Canada. SSRN Journal. Preprint posted online on April 6, 2020 2020. [doi: 10.2139/ssrn.3567474]
7. Xue J, Chen J, Hu R, Chen C, Zheng C, Liu X, et al. Twitter discussions and emotions about COVID-19 pandemic: a machine learning approach. ArXiv. Preprint posted online on May 26, 2020 2020 [FREE Full text]
8. Aslam S. Omnicore. URL: https://www.omnicoreagency.com/youtube-statistics [accessed 2021-02-11]
9. Basch CH, Basch CE, Ruggles KV, Hammond R. Coverage of the Ebola virus disease epidemic on YouTube. Disaster Med Public Health Prep 2015 Oct 19;9(5):531-535. [doi: 10.1017/dmp.2015.77] [Medline: 26088275]
10. Bora K, Das D, Barman B, Borah P. Are internet videos useful sources of information during global public health emergencies? A case study of YouTube videos during the 2015-16 Zika virus pandemic. Pathog Glob Health 2018 Sep;112(6):320-328 [FREE Full text] [doi: 10.1080/20477724.2018.1507784] [Medline: 30156974]
11. Pathak R, Poudel DR, Karmacharya P, Pathak A, Aryal MR, Mahmood M, et al. YouTube as a source of information on Ebola virus disease. N Am J Med Sci 2015 Jul;7(7):306-309 [FREE Full text] [doi: 10.4103/1947-2714.161244] [Medline: 26258077]
12. Khatri P, Singh SR, Belani NK, Yeong YL, Lohan R, Lim YW, et al. YouTube as source of information on 2019 novel coronavirus outbreak: a cross sectional study of English and Mandarin content. Travel Med Infect Dis 2020 Jul;35:101636 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101636] [Medline: 32205267]
13. COVID-19: Trudeau's daily briefing. YouTube. URL: https://www.youtube.com/playlist?list=PLeyJPHbRnGaYxLybblXjMbgiPdt6hhO7U [accessed 2021-02-11]
14. Thant W, Khaing S, Mon E. Preprocessing of YouTube Myanmar music comments for sentiment analysis. 2020 Presented at: 2020 IEEE Conference on Computer Applications (ICCA); February 27-28, 2020; Yangon, Myanmar p. 1-6. [doi: 10.1109/icca49400.2020.9022834]
15. Hucka M. Nostril: a nonsense string evaluator written in Python. JOSS 2018 May;3(25):596. [doi: 10.21105/joss.00596]
16. Ghag K, Shah K. Comparative analysis of effect of stopwords removal on sentiment classification. 2015 Presented at: International Conference on Computer, Communication and Control (IC4); September 10-12, 2015; Indore, India p. 1-6. [doi: 10.1109/ic4.2015.7375527]
17. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003 Jan;3:993-1022 [FREE Full text]
18. Obadimu A, Mead E, Hussain M, Agarwal N. Identifying toxicity within YouTube video comment. In: Social, Cultural, and Behavioral Modeling Lecture Notes in Computer Science. 2019 Presented at: 2019 International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation; July 9-12, 2019; Washington, DC p. 214-223. [doi: 10.1007/978-3-030-21741-9_22]
19. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
20. Jacobi C, van Atteveldt W, Welbers K. Quantitative analysis of large amounts of journalistic texts using topic modelling. Digit Journal 2015 Oct 13;4(1):89-106. [doi: 10.1080/21670811.2015.1093271]
21. Harris K. Canada to turn back asylum seekers, close border at midnight to stop spread of COVID-19. CBC News. 2020 Mar 20. URL: https://www.cbc.ca/news/politics/trudeau-covid19-coronavirus-medical-equipment-1.5504149 [accessed 2021-02-11]

22. Young N, Pebody R, Smith G, Olowokure B, Shankar G, Hoschler K, et al. International flight-related transmission of pandemic influenza A(H1N1)pdm09: an historical cohort study of the first identified cases in the United Kingdom. Influenza Other Respir Viruses 2014 Jan 07;8(1):66-73 [FREE Full text] [doi: 10.1111/irv.12181] [Medline: 24373291]
23. Bialek S, Allen D, Alvarado-Ramy F. First confirmed cases of Middle East respiratory syndrome coronavirus (MERS-CoV) infection in the United States, updated information on the epidemiology of MERS-CoV infection, and guidance for the public, clinicians, and public health authorities—May 2014. MMWR Morb Mortal Wkly Rep 2014 Jun 27;63(25):554. [doi: 10.1111/ajt.12841]
24. Brown CM, Aranas AE, Benenson GA, Brunette G, Cetron M, Chen TH, Centers for Disease Control and Prevention (CDC). Airport exit and entry screening for Ebola—August-November 10, 2014. MMWR Morb Mortal Wkly Rep 2014 Dec 12;63(49):1163-1167 [FREE Full text] [Medline: 25503920]
25. Gintova M. Social media use by government in Canada: examining interactions of immigration, refugees and Citizenship Canada on Twitter and Facebook. In: #SMSociety17: Proceedings of the 8th International Conference on Social Media & Society. 2017 Jul Presented at: 8th International Conference on Social Media & Society; July 2017; Toronto, ON p. 1-5. [doi: 10.1145/3097286.3097321]
26. Gintova M. Use of social media in Canadian public administration: opportunities and barriers. Can Public Admin 2018 Dec 18;62(1):7-26. [doi: 10.1111/capa.12298]
27. Bertot JC, Jaeger PT, Munson S, Glaisyer T. Social media technology and government transparency. Computer 2010 Nov;43(11):53-59. [doi: 10.1109/mc.2010.325]
28. SteelFisher GK, Blendon RJ, Bekheit MM, Lubell K. The public's response to the 2009 H1N1 influenza pandemic. N Engl J Med 2010 Jun 03;362(22):e65. [doi: 10.1056/NEJMp1005102] [Medline: 20484390]
29. Lemieux T, Milligan K, Schirle T, Skuterud M. Initial impacts of the COVID-19 pandemic on the Canadian labour market. Can Public Policy 2020 Jul 01;46(s1):S55-S65. [doi: 10.3138/cpp.2020-049]
30. Distributions of household economic accounts for income, consumption, saving and wealth of Canadian households, 2018. Statistics Canada. 2019 Mar 27. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/190327/dq190327b-eng.htm [accessed 2020-10-15]
31. Nearly 5.4 million receiving emergency federal aid. CTV News. 2020 Apr 13. URL: https://www.ctvnews.ca/politics/nearly-5-4-million-receiving-emergency-federal-aid-1.4893773 [accessed 2020-10-15]
32. Johnson R, Coleman R, Batten N, Hallsworth D, Spencer E. The quiet crisis of PhDs and COVID-19: reaching the financial tipping point. Research Square. Preprint posted online on June 18, 2020 2020. [doi: 10.21203/rs.3.rs-36330/v1]
33. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): a review. Int J Surg 2020 Jun;78:185-193 [FREE Full text] [doi: 10.1016/j.ijsu.2020.04.018] [Medline: 32305533]
34. Debnath R, Bardhan R. India nudges to contain COVID-19 pandemic: a reactive public policy analysis using machine-learning based topic modelling. PLoS One 2020 Sep 11;15(9):e0238972 [FREE Full text] [doi: 10.1371/journal.pone.0238972] [Medline: 32915899]

Abbreviations

ASCII: American Standard Code for Information Interchange
CBC: Canadian Broadcasting Corporation
CERB: Canada Emergency Response Benefit
CESB: Canada Emergency Student Benefit
LDA: latent Dirichlet allocation

Edited by G Eysenbach, C Basch; submitted 30.08.20; peer-reviewed by J Zhang, J Chen, R Guo, L Seuren; comments to author 21.09.20; revised version received 15.10.20; accepted 04.02.21; published 23.02.21.

Please cite as:
Zheng C, Xue J, Sun Y, Zhu T
Public Opinions and Concerns Regarding the Canadian Prime Minister's Daily COVID-19 Briefing: Longitudinal Study of YouTube Comments Using Machine Learning Techniques
J Med Internet Res 2021;23(2):e23957
URL: https://www.jmir.org/2021/2/e23957
doi:10.2196/23957
PMID:33544690

Original Paper

Features Constituting Actionable COVID-19 Dashboards: Descriptive Assessment and Expert Appraisal of 158 Public Web-Based COVID-19 Dashboards

Damir Ivanković[1*], MD, MBA; Erica Barbazza[1*], MSc; Véronique Bos[1], MA; Óscar Brito Fernandes[1,2], MSc, MEd; Kendall Jamieson Gilmore[3], MSc; Tessa Jansen[1], MPH, PhD; Pinar Kara[4,5], MSc; Nicolas Larrain[6,7], MA; Shan Lu[8], PhD; Bernardo Meza-Torres[9,10], MSc, MD; Joko Mulyanto[1,11], MSc, MD, PhD; Mircha Poldrugovac[1], MSc, MD; Alexandru Rotar[1], MSc, PhD; Sophie Wang[6,7], MPH; Claire Willmington[3], MSc; Yuanhang Yang[4,5], MSc; Zhamin Yelgezekova[12], MSc; Sara Allin[13], MSc, PhD; Niek Klazinga[1], MD, PhD; Dionne Kringos[1], MSc, PhD

[1]Department of Public and Occupational Health, Amsterdam UMC, Amsterdam Public Health Research Institute, University of Amsterdam, Amsterdam, Netherlands

[2]Department of Health Economics, Corvinus University of Budapest, Budapest, Hungary

[3]Laboratorio Management e Sanità, Institute of Management and Department EMbeDS, Scuola Superiore Sant'Anna, Pisa, Italy

[4]Danish Center for Clinical Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

[5]Department of Psychiatry, Aalborg University Hospital, Aalborg, Denmark

[6]OptiMedis AG, Hamburg, Germany

[7]Hamburg Center for Health Economics, University of Hamburg, Hamburg, Germany

[8]School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[9]Department of Clinical and Experimental Medicine, University of Surrey, Surrey, United Kingdom

[10]Nuffield Department of Primary Care and Health Services, University of Oxford, Oxford, United Kingdom

[11]Department of Public Health and Community Medicine, Faculty of Medicine, Universitas Jenderal Soedirman, Purwokerto, Indonesia

[12]Independent Researcher, Minneapolis, MN, United States

[13]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Damir Ivanković, MD, MBA
Department of Public and Occupational Health
Amsterdam UMC
Amsterdam Public Health Research Institute, University of Amsterdam
Meibergdreef 9
Amsterdam, 1105AZ
Netherlands
Phone: 31 981801700
Email: d.ivankovic@amsterdamumc.nl

Abstract

Background: Since the outbreak of COVID-19, the development of dashboards as dynamic, visual tools for communicating COVID-19 data has surged worldwide. Dashboards can inform decision-making and support behavior change. To do so, they must be actionable. The features that constitute an actionable dashboard in the context of the COVID-19 pandemic have not been rigorously assessed.

Objective: The aim of this study is to explore the characteristics of public web-based COVID-19 dashboards by assessing their purpose and users ("why"), content and data ("what"), and analyses and displays ("how" they communicate COVID-19 data), and ultimately to appraise the common features of highly actionable dashboards.

Methods: We conducted a descriptive assessment and scoring using nominal group technique with an international panel of experts (n=17) on a global sample of COVID-19 dashboards in July 2020. The sequence of steps included multimethod sampling of dashboards; development and piloting of an assessment tool; data extraction and an initial round of actionability scoring; a

workshop based on a preliminary analysis of the results; and reconsideration of actionability scores followed by joint determination of common features of highly actionable dashboards. We used descriptive statistics and thematic analysis to explore the findings by research question.

Results: A total of 158 dashboards from 53 countries were assessed. Dashboards were predominately developed by government authorities (100/158, 63.0%) and were national (93/158, 58.9%) in scope. We found that only 20 of the 158 dashboards (12.7%) stated both their primary purpose and intended audience. Nearly all dashboards reported epidemiological indicators (155/158, 98.1%), followed by health system management indicators (85/158, 53.8%), whereas indicators on social and economic impact and behavioral insights were the least reported (7/158, 4.4% and 2/158, 1.3%, respectively). Approximately a quarter of the dashboards (39/158, 24.7%) did not report their data sources. The dashboards predominately reported time trends and disaggregated data by two geographic levels and by age and sex. The dashboards used an average of 2.2 types of displays (SD 0.86); these were mostly graphs and maps, followed by tables. To support data interpretation, color-coding was common (93/158, 89.4%), although only one-fifth of the dashboards (31/158, 19.6%) included text explaining the quality and meaning of the data. In total, 20/158 dashboards (12.7%) were appraised as highly actionable, and seven common features were identified between them. Actionable COVID-19 dashboards (1) know their audience and information needs; (2) manage the type, volume, and flow of displayed information; (3) report data sources and methods clearly; (4) link time trends to policy decisions; (5) provide data that are "close to home"; (6) break down the population into relevant subgroups; and (7) use storytelling and visual cues.

Conclusions: COVID-19 dashboards are diverse in the why, what, and how by which they communicate insights on the pandemic and support data-driven decision-making. To leverage their full potential, dashboard developers should consider adopting the seven actionability features identified.

(J Med Internet Res 2021;23(2):e25682) doi:10.2196/25682

KEYWORDS

COVID-19; pandemic; internet; performance measures; public reporting of health care data; public health; surveillance; health information management; dashboard; accessibility; online tool; communication; feature; expert

Introduction

Since the outbreak of COVID-19, public reporting of pandemic-related indicators such as new cases, death counts, and testing rates has surged. This heightened level of activity attests to the core function of governments to protect the public's health and safety as well as their critical role of providing information to achieve this end [1-4]. The uses and advantages of publicly reporting health information are known. They include enabling international comparisons [5,6]; monitoring and improving the quality of care [1,6,7]; fostering accountability and transparency [8-10]; empowering the public to form an opinion on and build trust in their government's response; and supporting individuals to make informed, risk-minimizing behavior changes [11,12].

Dashboards are a dynamic modality for reporting data visually; they are typically designed as a single screen with the aim of quickly and effectively presenting users with critical information to act upon [13-15]. Unlike static reporting modalities, such as articles or reports, dashboards have the potential to present real-time (or near–real-time) data updates at a glance [15]. In the health sector, dashboards have been relied on for health system performance assessments [15,16], internal management [17,18], and responses to earlier outbreaks [19,20].

In 2020, the urgent worldwide need for COVID-19 data, coupled with the penetration of the internet [21], digitalization of health information systems [22,23], and access to open-source web-based software [24], has enabled unmatched speed, scale, and diversification of actors in the development of dashboards to monitor and report on the COVID-19 pandemic. As a result, public web-based dashboards have been widely adopted as a reporting modality for COVID-19 data. Examples extend well beyond national, regional, and local governments to include dashboards by international organizations (eg, the World Health Organization (WHO) [25]), academia (eg, the Johns Hopkins Coronavirus Resource Center [26,27]), and industry (eg, Deloitte [28]), as well as independent initiatives (eg, nCoV2019.live [29]).

Although COVID-19 dashboards may be widely accessible, their effective *use* to modify the course of the pandemic through the translation of data to information, information to opinions, and opinions to decision-making is determined by their actionability. To be actionable, the information should be both *fit for purpose*—meeting a specific information need—and *fit for use*—placing the right information into the right hands at the right time and in a manner that can be understood [30-32]. In other words, the mere accessibility of COVID-19 dashboards does not guarantee data-informed decision-making [12,33]. Although communication sciences, health promotion, and the emerging field of health care performance intelligence offer insights into the effective delivery of information [14,33-36], the factors that make dashboards actionable in the context of COVID-19 have yet to be rigorously assessed.

In this study, we set out to explore the state of the art of publicly available web-based COVID-19 dashboards and identify the features conducive to their actionability. To do so, we took a "snapshot" of this dynamic landscape and assessed COVID-19 dashboards in July 2020. The resulting overview of the dashboard landscape served both to take stock of their use in this initial period and to accelerate their progress in the phases still to come. With these aims, the study was guided by four key questions: (1) Why and for whom were COVID-19 dashboards developed? (2) What information do they provide?

(3) How is this information analyzed and presented? and (4) What are the common features of highly actionable dashboards?

Methods

Study Design

We conducted an observational descriptive assessment and scoring using nominal group technique (NGT) [37,38] on a global sample of COVID-19 dashboards. Each dashboard was reviewed using a study-specific assessment tool that was piloted and validated among a panel of scorers (n=17) prior to its use [37,38]. NGT was chosen over other consensus methods (eg, Delphi) for scorers to independently appraise a subset of dashboards using the assessment tool and collectively discuss what makes them actionable through a series of workshops [38,39]. All workshops were conducted virtually rather than face-to-face in accordance with pandemic-related public health measures.

Panel of Scorers

A panel of scorers was assembled through an existing international network of health care performance intelligence researchers [40]. The scorers had common expertise and training in health care performance data and the use of these data for management and governance. Collectively, the scorers (8 women, 9 men) were of 15 nationalities and were proficient in more than 20 languages (Bosnian, Catalan, Chinese, Croatian, Danish, Dutch, English, French, German, Indonesian, Italian, Kazakh, Malay, Montenegrin, Norwegian, Portuguese, Romanian, Russian, Serbian, Slovenian, Spanish, Swedish, and Turkish). This enabled the dashboards to be assessed in their original languages rather than through translations, avoiding the use of translation software and its limitations when used with data visualizations.

Inclusion and Exclusion Criteria

We defined a COVID-19 dashboard based on the following criteria: (1) reporting of key performance indicators related to the COVID-19 pandemic; (2) the use of some form of data visualization; (3) dynamic reporting, meaning the data are updated regularly; and (4) public availability in a web-based format. No restrictions were placed on a dashboard's primary level (eg, international, national, regional, or local) or the type of organization responsible for its development (eg, international, governmental, academia, news or media, industry, or private initiative). We excluded dashboards that were available only via mobile apps (eg, Telegram) or that required users to log in (eg, Facebook). Dashboards beyond the language competencies of the panel of scorers were also excluded.

Step One: Dashboard Sampling

Our search strategy for dashboards aimed to be thorough but not exhaustive. This was in line with our aim of exploring the state of the art of public web-based COVID-19 dashboards. An initial list of dashboards was collected through sampling conducted from May 19 to June 30, 2020. Three methods were applied: (1) surveying the authors; (2) surveying other international networks of public health, health services, and system researchers and practitioners (Young Forum Gastein, European Public Health Association, and European Network of Medical Residents in Public Health); and (3) snowballing of sources identified through (1) and (2). The sampling survey was developed using a Google Forms data collection tool and disseminated by email (Multimedia Appendix 1).

The consolidated list of dashboards was screened by one team member with the aims to confirm the inclusion criteria were met; exclude duplicates; and assess the available languages for each dashboard against the panel's competencies. Dashboards were labeled as red (exclude), green (include), or yellow (obtain second opinion). A second team member assessed dashboards labeled yellow, from which a final joint decision on inclusion or exclusion was made.

Step Two: Developing an Assessment Tool

An assessment tool was developed by drawing primarily on two existing theoretical models. From communication sciences, we applied the Lasswell model (1948) [41], which states that for mass communication processes to be understood, each element of "who (says) what (to) whom (in) which channel (with) what effect" has to be presented and understood. These five elements—the communicator, message, medium, audience, and effect—informed the basis of the assessment tool's considerations. We tailored these considerations to the communication of COVID-19 data by drawing on the emerging discipline of performance intelligence in health [36,42]. Specifically, we incorporated key considerations from a definition of actionability and its notions of fitness for purpose and use (Barbazza et al, unpublished data, 2021). The resulting considerations are in line with existing health information instruments (eg, [43,44]), although they were tailored to the aims of the study.

These considerations were clustered to depict COVID-19 dashboards by their general characteristics and a description of why, what, and how data is communicated, followed by an appraisal of their overall actionability (Table 1). Actionability scores were defined on a Likert scale from "not actionable" (score=1) to "extremely actionable" (score=5) and assigned based on the scorer's judgement of the considerations assessed and their expert opinion of the dashboard's fitness for purpose and use. Scores were accompanied by a written statement explaining the rationale behind the response. In line with the study's aim to consolidate key features of highly actionable dashboards, the scoring was merely a means to this end: the panel's individual appraisal of actionability facilitated the clustering of the actionability of the dashboards as low (score=1 or 2) or high (score=4 or 5) for further collective deliberation on their common features.

Table 1. Overview of the assessment tool.

Cluster	Considerations
General characteristics	• Level (scale) of focus • Responsible organization and type • Languages available • Scope of web page information
Why	• Purpose of use of the dashboard • Intended audience (user)
What	• Indicator titles • Data sources • Availability of metadata • Frequency of data updates
How	• Use of time trend for analysis • Geographic level (scale) of analysis • Types of possible breakdowns • Use of visualizations • Degree of interactivity • Use of simplicity techniques
Actionability score	• Overall appraisal of actionability

An Excel-based tool (Microsoft Corporation) was developed to record our findings. Each consideration of the assessment tool was formulated as a question with defined answer options. The tool included the underlying theory for the considerations by referring back to the concepts applied and available evidence [1,2,5,16,30,31,33,45-55] (Multimedia Appendix 2) to remind the panel of the significance of each consideration and aid the assessment and scoring process.

Step Three: Piloting and Calibrating

A prototype of the assessment tool was piloted by two authors on five dashboards. The extracted data were reviewed jointly with two other team members. This resulted in refinements to the phrasing of the questions and answer options. A second iteration of the assessment tool was then piloted with the panel of scorers on a sample of 18 dashboards representing a range of contexts, levels, and organization types. Each dashboard was independently reviewed by two scorers. Prior to piloting, a virtual training session with the panel of scorers was organized, recorded, and disseminated to serve as a resource. Each scorer was given six days (June 17-22, 2020) to review their two assigned pilot dashboards.

The pilot data were reviewed to assess the consistency of responses (ie, scorers of the same dashboard recorded equivalent answers) and meaningfulness of the answers (ie, the answer categories were meaningfully differentiated between dashboards). Where possible, the open-ended answer options of the tool were further specified into categorical values based on recurrent themes in the pilot data set. Definitions were added for key terms based on comments by the scorers. The reviewed pilots and tool amendments were returned to the panel of scorers, and a follow-up meeting was organized to discuss the reviews.

Step Four: Data Extraction and Round One Scoring

Each scorer was assigned between 5 and 12 dashboards to assess. The dashboards were distributed with first order priority given to the language competencies of each scorer. To synchronize the assessment, the scorers were given a 2-week period to complete data extraction. The assessment was limited to each dashboard's main page, and a "one-click-away policy" was applied by which content accessible within one click of the main page was also assessed. To store a record of the dashboard on the date it was reviewed, the main page of each dashboard was archived, generating a permanent and publicly available record of its contents [56].

Step Five: Preliminary Analysis and First Consensus Workshop

The data records from each scorer were consolidated by the lead authors into a master data set for analysis and subsequently underwent a series of data quality checks to detect data entry errors, inconsistences, or missed fields. In all instances where errors were detected, corrections were suggested and discussed jointly; once agreed upon, the changes were entered into the master data set.

The findings were totaled and averaged by research question. Free text fields and comments were analyzed in a deductive and inductive approach: topics explored in the tool (Multimedia Appendix 2) were used to guide the deductive thematic analysis [57], and new themes that emerged were identified using an inductive approach [58]. This included an analysis of indicator titles using an existing classification of types of pandemic-related information [3]. Due to the observed variability in phrasing of indicator titles and calculations, key performance indicators were grouped by themes.

A workshop with the panel of scorers was organized to discuss the preliminary results and distribution of actionability scores. During the workshop, panelists individually shared the rationale for their scoring of dashboards with low (score=1 or 2) and high (score=4 or 5) actionability. The common features of dashboards scored as highly actionable were discussed to further calibrate

the panel's scoring of the actionability. From this discussion, a working list of actionability features was consolidated.

Step Six: Round Two Scoring and Second Consensus Workshop

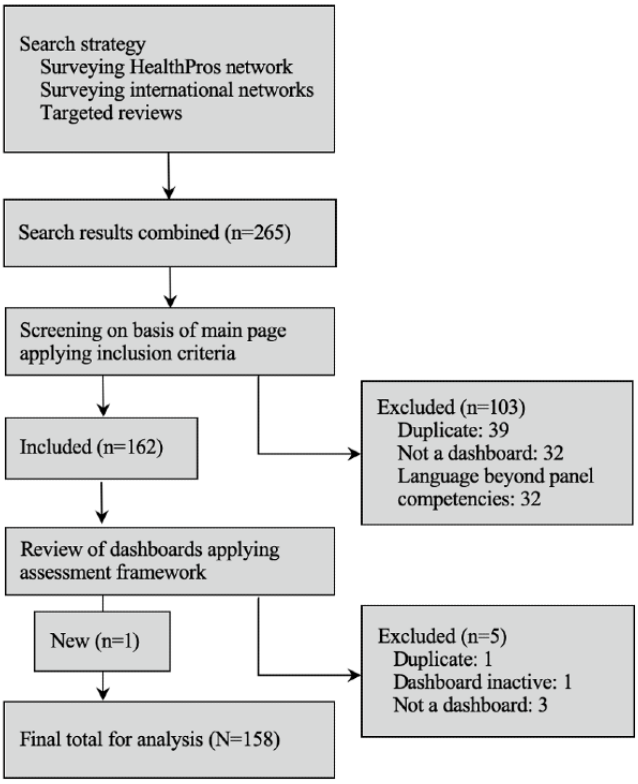
All panelists were returned their original data records and given 1 week to revisit their initial actionability scoring, drawing on the discussion during the workshop. Panelists were given the opportunity to increase each score, lower it, or leave it the same. Following rescoring, the distributions of the scores were recalculated. The data records for the top dashboards (score=5) following this second round were consolidated and provided to the panel, together with the working set of actionability features. A second consensus workshop was convened and, in a similar way to the previous workshop, a round table was conducted for each scorer to share their views. This was followed by a joint discussion to reach agreement on the common features of highly actionable dashboards.

Results

Identified Dashboards

Using our multimethod search strategy, we initially identified 265 COVID-19 dashboards. More than 40 respondents contributed to the sampling survey, including all members of the study team and international public health experts. Following screening of each dashboard's main page, 103 dashboards were excluded. The remaining 162 dashboards were distributed among the panel of scorers for full review. During the assessment process, 5 additional dashboards were excluded and 1 new dashboard was included. A final total of 158 dashboards was included for further analysis (Figure 1).

Figure 1. Flow diagram of COVID-19 dashboard sampling.



Data extraction and the first round of scoring were conducted in a 2.5-week period between July 6 and 23, 2020. The data extract and archived version of each dashboard were referred to throughout the study. Therefore, any updates following this date were not accounted for. The 158 dashboards were assessed in 22 different languages, predominately in English (n=85, 53.8%), followed by Russian (n=11, 7.0%), Spanish (n=9, 5.7%), French (n=9, 5.7%), and Chinese (n=6, 3.8%). A full listing of the dashboards assessed is available in Multimedia Appendix 3.

General Description of the Assessed COVID-19 Dashboards

Table 2 summarizes key characteristics of the 158 dashboards assessed. Our sample included dashboards reporting on 53 countries in all 6 WHO regions [59]. On the date of the review, the severity of the pandemic with regard to total cases and deaths varied widely between location as reported in Multimedia Appendix 3.

More than half of the dashboards (93/158, 58.9%) were developed for use at the national level. Nearly two-thirds of the dashboards (100/158, 63.3%) were developed by government authorities at the national, regional, or municipal level. New initiatives or organizations formed in response to COVID-19 accounted for 10.1% (16/158) of the dashboards assessed [29,60-74].

With regard to language, only one-fifth of the dashboards were available in more than one language with full functionality (32/158, 20.3%). In terms of their scope of information, gauged according to the content of the dashboard as well as information to which users were redirected through affiliate links, almost

all the dashboards were epidemiological in focus (156/158, 98.7%), followed by providing information on infection control measures and health system management (65/158, 41.1%, and 49/158, 31.0%, respectively).

Table 2. Characteristics of the assessed COVID-19 dashboards (N=158) from 53 countries.

Characteristic	Value, n (%)
Region[a]	
Global	20 (12.7)
Europe and Central Asia	63 (39.9)
North and South America	45 (28.5)
Western Pacific	22 (13.9)
Southeast Asia	4 (2.5)
Africa	3 (1.9)
Eastern Mediterranean	1 (0.6)
Level	
International	25 (15.8)
National	93 (58.9)
Regional (provincial, state, county)	33 (20.9)
Municipal (city, district)	7 (4.4)
Type of organization	
International organization	7 (4.4)
Governmental	100 (63.3)
Academia	9 (5.7)
News or media outlet	14 (8.9)
Industry	9 (5.7)
Independent initiative	16 (10.1)
Other	3 (1.9)
Languages available with full functionality[b]	
One language	126 (79.7)
Two languages	22 (13.9)
Three or more languages	10 (6.3)
Additional languages available with reduced functionality[c]	
One or more languages	16 (10.1)
Scope of information[d]	
Epidemiological information	156 (98.7)
Infection control measures	65 (41.1)
Health system management	49 (31.0)
Social and economic implications	31 (19.6)
Population behavioral insights	25 (15.8)
Other	28 (17.7)

[a]Country status and region according to the WHO classification [59].

[b]Full functionality: the webpage is equivalent in the different languages.

[c]Reduced functionality: the webpage is available in additional languages but with less information and fewer functionalities compared to the main languages.

[d]According to the WHO classification [3].

Uses and Users of COVID-19 Dashboards

A quarter of the dashboards (45/158, 28.5%) explicitly stated the intended purpose of their reporting. Of these 45 dashboards, the statements spanned three main themes: (1) high-level reporting to create trust and aid overall compliance (25/45, 56%); (2) sub-national reporting targeting policy interventions, including benchmarking (12/45, 27%); and (3) individual-risk assessment (8/45, 18%).

Only 14.6% (23/158) of the dashboards explicitly stated the intended audience (end users). Target users predominately included the general public (20/23, 87%) and, in a few instances, more specific audiences such as travelers or subject matter experts (6/23, 26%). When examined by the level of reporting, national-level dashboards were less likely to explicitly state the intended audience (9/93, 10%), while international- and municipal-level dashboards were more likely to do so (7/25, 28%, and 2/7, 29%, respectively).

Of the 158 dashboards assessed, 20 (12.7%) reported both the purpose and intended user explicitly. The profiles of these dashboards, in terms of their levels of reporting and the types of organizations that developed them, did not differ from the characteristics of the general sample. For the remainder of the analysis, the sample of dashboards was aggregated rather than subdivided by the intended purpose of the use and audience, due to the limited explicit statements of both.

Content and Data of COVID-19 Dashboards

Key Performance Indicators

Table 3 summarizes the frequency of indicator themes reported by the dashboards. See Multimedia Appendix 4 for illustrative examples of indicator titles. On average, the dashboards reported on 5.3 indicator themes (maximum 15, minimum 1). Almost all the dashboards reported public health and epidemiological indicators (155/158, 98.1%), particularly those that reported on cases and deaths. These account for the only high-frequency indicator themes (indicators present in more than two-thirds of the assessed dashboards). Medium-frequency indicator themes (themes reported in more than one-third but less than two-thirds of dashboards) were related to hospital care (hospitalizations, admissions to infection control units), testing (total tests, testing rates), and spread and death (recovered and active cases).

Only 4.4% of the dashboards (7/158) reported indicators related to social and economic impacts. Indicator themes included employment and hardship relief (eg, [28,75]) and transport, trade, and international travel (eg, [28,75]). Indicators of behavioral insights were also infrequently reported (8/158, 5.1%). Indicator themes included two main types: (1) self-reported adherence related to restrictions (eg, [76,77]) or health and well-being status (eg, [75]) and (2) observed public adherence to restrictions assessed through mobility data or reported breaches of restrictions (eg, [60,78]).

Some use of composite scores to signal overall risk levels or the current status by sector (eg, health, economy) was identified, although this use was infrequent (eg, [28,61,79]).

Table 3. Frequency of indicator themes reported for the 158 dashboards assessed.

Information type and cluster	Indicator themes	Value, n (%)	Frequency[a]
Public health and epidemiological			
Spread and death	Cases (all confirmed cases)	150 (94.9)	High
	Deaths	136 (86.1)	High
	Recovered (healed, cured)	91 (57.6)	Medium
	Active cases	56 (35.4)	Medium
	Mortality rate (case fatality rate)	24 (15.2)	Low
	Reproduction rates (attack rate)	12 (7.6)	Low
	Future projections/risk models	5 (3.2)	Low
	Doubling rate	3 (2.0)	Low
Testing	Testing (total number tested, PCR[b] tests)	80 (50.6)	Medium
	Testing rates (positivity, negative tests)	43 (27.2)	Medium
	Tests–pending results	17 (10.8)	Low
	COVID-19 antibody tests (serology tests)	1 (0.6)	Low
Risk management	Self-quarantine (isolation notices)	18 (11.4)	Low
	Contact tracing	6 (3.8)	Low
Health system management			
Hospital care	Hospitalized (admissions, discharges)	74 (46.8)	Medium
	Admitted to ICU[c] (critical condition)	47 (29.7)	Medium
	On a ventilator	14 (8.8)	Low
Health system capacity	Hospital bed capacity (availability)	12 (7.6)	Low
	ICU bed capacity	10 (6.3)	Low
	Ventilator capacity (available ventilators)	5 (3.2)	Low
	Non–COVID-19 service usage	4 (2.5)	Low
	Personal protective equipment stock	2 (1.3)	Low
	Testing stock	2 (1.3)	Low
Social and economic impact			
N/A[d]	Employment and hardship relief	7 (4.4)	Low
	Transport, trade, and international travel	3 (1.9)	Low
Behavioral insights			
N/A	Observed public adherence to restrictions	4 (2.5)	Low
	Self-reported adherence to restrictions	2 (1.3)	Low
	Self-reported health and well-being status	2 (1.3)	Low

[a]Low: ≤33%; medium: 34%-66%; high: ≥67%.

[b]PCR: polymerase chain reaction.

[c]ICU: intensive care unit.

[d]N/A: not applicable.

Data Sources and Metadata

One quarter of the dashboards did not explicitly report the source of their data (39/158, 24.7%). National-, regional-, and municipal-level government-run dashboards predominately reported the use of data sourced from official public health authorities. International dashboards predominately reported the use of data sourced from the WHO [25] or the Johns Hopkins Centre for Systems Science and Engineering [26].

Less than half of the dashboards (63/158, 39.9%) specified metadata (data dictionaries, indicator specifications) in the format of notes, footnotes, or linked additional web pages to provide further information on the methodology by which an indicator was calculated. Of the 158 dashboards, 39 (24.7%)

did not report their data sources or metadata details. The majority of dashboards updated their data daily and explicitly stated the update frequency and time of the last update.

Types of Analysis and Presentation of Data on COVID-19 Dashboards

Table 4 summarizes the types of analysis and presentation of data. The dashboards predominately reported indicators over time (138/158, 87.4%), and most of these breakdowns were by day (128/138, 92.8%). Of the dashboards, 40% reported data on two geographic levels (eg, national and regional or regional and municipal). In the case of national-level dashboards (n=93), geographic breakdowns predominately included regional comparisons (73/93, 79%), with some municipal-level (28/93, 30%) and international-level (25/93, 27%) comparisons. Breakdowns by neighborhood (post–code-level) were reported in only a few instances (4/93, 4%).

In addition to geographic breakdowns, more than half of the dashboards (96/158, 60.8%) analyzed data by other breakdowns: on average, three types of breakdowns were included. Of these 96 dashboards, the most common breakdowns included by age (79/96, 82%), sex (71/96, 74%), and mode of transmission (26/96, 27%). Other breakdowns, although less frequently reported, included race, ethnicity, long-term care facilities, health care workers, comorbidities, and socioeconomic status.

As per our inclusion criteria, all dashboards used some form of visualization. On average, two types of visualizations were included per dashboard. These included graphs or charts (134/158, 84.8%), maps (111/158, 70.3%), and tables (95/158, 60.1%). Almost half of the dashboards (76/158, 48.1%) did not include written descriptions to clarify either the quality or meaning of the data, while 31/158 dashboards (19.6%) provided both.

More than half of the dashboards (104/158, 65.8%) used some technique to simplify the data. In these 104 dashboards, color-coding was most often used (n=93, 89.4%), followed by size variation (n=40, 38.5%). The majority of dashboards (126/158, 79.7%) included some element of user interaction. These elements mostly included the possibility to present more information (eg, pop-up windows), change the information (eg, different breakdowns), or change the display (eg, switch from table to map).

Table 4. Summary of analysis and presentation of dashboard information.

Considerations	Value, n (%)
Time trend analysis availability (N=158)	
Time trend analysis available	138 (87.3)
No time trend analysis	20 (12.7)
Use of time trend analysis (n=138)[a,b]	
By day	128 (92.8)
By week	33 (23.9)
By month	19 (13.8)
Geographic levels (scales) of analysis (N=158)[b]	
International (multicountry)	54 (34.2)
National	118 (74.7)
Regional	117 (74.1)
Municipal	54 (34.2)
Neighborhood	13 (8.2)
Other	5 (3.2)
Number of levels (scales) of analysis per dashboard (N=158)	
1 level	34 (21.5)
2 levels	65 (41.1)
3 or more levels	59 (37.3)
Disaggregation availability per dashboard (N=158)	
1 or 2 types of disaggregation	48 (30.4)
3 or 4 types of disaggregation	42 (26.6)
5 or more types of disaggregation	6 (3.8)
No disaggregation options	62 (39.2)
Disaggregation options (n=96)[a,b]	
Age	79 (82.3)
Sex	71 (74.0)
Mode of transmission	26 (27.1)
Long-term care facilities	16 (16.7)
Ethnicity	12 (12.5)
Race	10 (10.4)
Health workers	9 (9.4)
Comorbidities	9 (9.4)
Socio-economic status	2 (2.1)
Other	23 (24.0)
Visualization features (N=158)[b]	
Graphs/charts	134 (84.8)
Maps	111 (70.3)
Tables	95 (60.1)
Video/animations	10 (6.3)
Use of narratives to interpret data (N=158)	
Yes, to clarify the quality of the data only	28 (17.7)

Considerations	Value, n (%)
Yes, to clarify the meaning of the data only	23 (14.6)
Yes, to clarify both the quality and the meaning	31 (19.6)
No	76 (48.1)
Simplification techniques used (n=104)[a,b]	
Use of color-coding	93 (89.4)
Size variation	40 (38.5)
Icons	6 (5.8)
Interactive options (n=126)[a,b]	
More information	115 (91.3)
Change of information	61 (48.4)
Change of display	44 (34.9)

[a]Subset of applicable dashboards (ie, 138 dashboards that *do* use time trends).

[b]Percentages for these considerations do not total to 100%, as multiple considerations could be present per dashboard.

Features of Actionable Dashboards

In the first round of scoring, 21 of the 158 dashboards assessed (13.3%) were scored with the highest actionability score (score=5), and 18 dashboards (11.4%) received the lowest score (score=1), for a mean score of 3.01 (SD 1.20). The second round of scoring resulted in a final total of 20 dashboards that were scored as the most actionable. A quarter of the dashboards (40/158, 25.3%) were scored differently: 24 scored lower, and 16 scored higher. All 17 panelists completed both rounds of scoring. Details on the distribution of scoring by panelist and between rounds are summarized in Multimedia Appendix 5.

The panel workshop following the first round of scoring resulted in a total of 18 features that characterized highly actionable dashboards. After rescoring, these features were further discussed among the panel to consolidate the list in terms of their description and importance as well as its consistency and completeness as a set. A final total of seven key features common to highly actionable dashboards were agreed upon (Table 5). There was consensus among the panelists that some dashboards excelled in certain features over others. These dashboards are noted as illustrative examples.

Table 5. Seven features of highly actionable COVID-19 dashboards.

Number	Feature	Explanation	Examples
1	Know the audience and their information needs	Dashboards with a known audience and explicit aim had focus and continuity in their content, analysis and delivery. Techniques such as guiding key questions or overall composite scores clearly communicated the decision they intended to support. Multilanguage functionality and exact timing of updating signaled an awareness and intent to encourage their regular use by the intended decision maker.	#HowsMyFlattening [60], Covid Act Now [61], State of California [79].
2	Manage the type, volume, and flow of information	The selection of a concise number of indicators brought focus and importance to the information and the possibility to view indicators together at a glance. The use of indicators in moderation, although still spanning varied types of information, was especially effective. The ordering of information, from general to specific or in sections based on theme, made the flow of information intuitive.	Covid Act Now [61] reports on five key indicators. Deloitte [28] and the City of Vancouver [78] included a range of types of information.
3	Make data sources and methods clear	A clear source of data and explanation of an indicator's construction, including potential limitations, was found to be an important component of trust in the dashboard and clarity in its reporting. This information can be provided in short narratives that support users to understand what is in fact being presented.	Denmark [80], France [76], Spain [81], and media pages of the Canadian Broadcasting Corporation [82] and the New York Times [83] paid attention to narrating the calculation of indicators.
4	Link time trends to policy (decisions)	Reporting data over time together with the introduction of key infection control measures facilitated an understanding of their effect (or lack thereof). This was found to be conducive to generating public support for infection control measures.	ABC News [84] and Sledilnik [62] embed policy measures over time. The City of Toronto [85] reports city targets.
5	Provide data "close to home"	To inform individuals of risks in their immediate surroundings, granular geographic breakdowns are needed. Data that are highly aggregated are difficult to understand. Maps (over tables and charts) were most effective to provide geographic information.	The United Kingdom [86] offers post–code-level breakdowns. Germany [87] provided city- and borough-level information for Berlin.
6	Break down the population to relevant subgroups	Providing data with the possibility to explore varied population characteristics made indicators relatable to individual users. It enables understanding of risks and trends based on one's own demographics. It can also facilitate equity-driven decision-making by exposing differences among the population.	Ethnicity and race breakdowns were provided in New Zealand [75] and various US dashboards [79,88-92]. #HowsMyFlattening [60] provided breakdowns on economic status.
7	Use storytelling and visual cues	A concise narrative explaining the significance of a trend supports users to understand the importance of the information. Bare statistics without a narrated analysis leave the burden of interpretation solely to the user. Brief explanations on the meaning of trends used in combination with visual techniques, such as intuitive color schemes and icons, supported ease of interpretation.	Covid Act Now [61] narrates the significance of trends. The State of Colorado [88] uses colored icons to signal the direction of trends.

Discussion

Principal Findings

With this study, we set out to assess the state of the art of public web-based COVID-19 dashboards globally during the initial stage of the pandemic (July 2020) and identify features common to the dashboards that were found to be highly actionable. We assessed 158 dashboards, each operating in a different context. Their differences aside, the dashboards analyzed in this study ultimately share a common aim: to serve as both a communication tool and call for individual and collective action to respond to the COVID-19 pandemic. Despite their contextual differences (or because of them), our results indicate that some dashboards fulfill their function of communicating, informing decision-making, and supporting behavior change better than others. Moreover, while it is also clear there is no single approach to developing a dashboard, our results suggest that introducing certain features may enhance the actionability of a dashboard.

Knowing the audience and their information needs was identified as a key actionability feature, which corresponds with the Lasswell model for effective communication ([1,41]; Barbazza et al, unpublished data, 2021). However, clear reporting of a dashboard's purpose (its "why") and audience (for "whom") was infrequent. This may be explained in part by the fact that the majority of the dashboards were developed by public authorities and hosted on existing web pages. Hence, the target audience (citizens) and the aim (constitutional mandate to protect health) may be considered implicit. However, without clarity on the intended use and user of a dashboard, its development is steered by the *potential* to be useful rather than addressing a specific information need [32,93-95].

"What" a dashboard communicates through its content is not a neutral window into the available data. It is the result of

judgment, discernment, and choice [14]. The average of 5 indicator themes reported per dashboard can be considered to be a manageable volume and is in line with the evidence that "less is more" [33,47]. It is the breadth of types of information presented that is concerningly narrow, with only a handful of dashboards addressing the WHO-recommended four types of information needed for a complete picture of the pandemic [3]. For example, indicators reporting on population behavioral insights gauge the compliance of citizens with infection control measures; thus, they are an important tool for maintaining public trust. However, in our sample, this type of information was rarely reported. This may be due to data infrastructure limitations and the limited availability of these data, especially in the early phases of the pandemic. Similarly, less than half of the dashboards reported on health system management indicators, despite the importance of these indicators in informing the management of both COVID-19 and non–COVID-19 services. Dashboards that did report on these non-epidemiological types of information may serve as inspiration for drawing on innovative data sources and indicators [28,60].

Clarity around data sources and indicator calculations (metadata) are critical for overall quality, credibility, and trustworthiness of reporting [46,48,49]. For transparency on how data were collected and insights into "what lies behind" the reported indicators, providing explicit data sources and calculations should be considered a minimum requirement. Nonetheless, our findings signal that these provisions are not a given. Further efforts are needed internationally and nationally to standardize indicator calculations and set requirements of what constitutes good practice in public reporting of pandemic-related data.

In terms of "how" content is presented, dashboards should be viewed as tools for making clear links between current trends and past policy decisions and individual behavior. Doing so connects change-points and actions, which has been found to contribute to an indicator's use [96,97]. It also serves to leverage the two-way communication potential of dashboards. Dashboards that fail to make the connection between the past and present miss the opportunity to communicate the effects of users' decision-making back to them. Beyond describing the past and present, only a handful of dashboards went further and employed predictive analytics by illustrating different future scenarios of "what could happen." The lack of precision of predictive models and simulations early in the pandemic likely stunted their use. Use of both descriptive and predictive approaches to dashboard design and tighter links between infection control policies and their effects should be further explored into the next phases of the pandemic.

We found frequent use of different display options and interactive techniques among the dashboards assessed. However, the analysis of data by location and by population subgroups was limited overall, which may restrict their utility for individual-level decision-making purposes. The challenge to report data locally and disaggregate the data by relevant breakdowns such as age, sex, socioeconomic status, and ethnic or racial groups may be in large part due to data infrastructure limitations and perceived legal obstacles [98]. Without collecting, registering, and using data about meaningful population subgroups, there is a risk of not being informed about these important (and modifiable) differences [98].

Finally, an actionable dashboard is based on complete, timely, and transparent data that is prepared, contextualized, and presented so that it can be used as information [99]. Our assessment found an overall underuse of known and proven delivery techniques, in particular, the use of explanatory narratives. Plain language text to clarify complicated information has proven to make end users more motivated and confident in using information in their decision-making [1,47,54]. Although commonly used software for the development of dashboards (eg, ArcGIS) has served to optimize their single-screen design, the embedding of narratives into templates may be useful for improving interpretation.

Future research could explore the following points. First, recognizing the highly dynamic nature of COVID-19 dashboards, a follow-up study could provide insights into how dashboards have evolved over time, given improvements in disease prevention, testing, and treatment as well as data infrastructure. Second, exploring across official municipal, regional, and national dashboards in a given context was beyond the scope of this study; however, such an exploration may offer insights into the possibility of tailoring dashboards at different levels to specific purposes and audiences. Third, this study has pursued a theoretically informed expert-based appraisal of actionability. A study from the perspective of the target audience is therefore merited and needed to obtain insights from firsthand use. Finally, the developed assessment tool could be used within a specific country context to analyze actions needed to implement the identified features.

Strengths and Limitations

To our knowledge, this is the most comprehensive summary of COVID-19 dashboards and assessment of their actionability published to date. The search for COVID-19 dashboards was wide-reaching and used multiple methods to amass a global sample. The approach tapped into a unique and highly specialized international network dedicated to health care performance intelligence, allowing for an expert, context-aware, and multicultural team. The multilanguage competencies of the panel made it possible for the dashboards to be reviewed in their original languages for high-quality data extraction. Through detailed data extraction and a structured process of scoring with joint deliberation, we have identified a set of timely and pragmatic features for optimizing dashboards. This is also the first study to our knowledge on the use of dashboards for public reporting from a communication and health care performance intelligence perspective. Importantly, the study was conducted at pace with the ongoing COVID-19 pandemic to ensure the potential for findings to inform the continued development of dashboards in combination with other communication tools.

We acknowledge the following potential limitations. First, the sample of dashboards is ultimately a subset of publicly available web-based COVID-19 reporting. The sample is also skewed to locations in the European and Pan-American regions, which account for two-thirds of the dashboards reviewed. This can be attributed in part to factors including the thorough but not exhaustive sampling strategy applied; the exclusion of

dashboards beyond the 22 language competencies of the panel (ie, Arabic and Hindi); and the focus on web-based dashboards to the exclusion of those exclusively on mobile apps (common to Asian countries). As an exploratory study, reasonable diversity of locations, in combination with different levels (scales) of focus and types of organizations, took precedent and was achieved. Nonetheless, the findings may not be generalizable to all contexts. Second, despite our best efforts to obtain a snapshot of COVID-19 dashboards in a common 2-week period, the severity and specific phase of the pandemic inevitably varied greatly on the date of the review as described. Our approach to assess rather than evaluate the impact of COVID-19 dashboards mitigates the significance of these differences on our findings. Third, the appraised actionability of the dashboards ultimately does not confirm their use in practice, and evaluating this was beyond the scope of this study.

Conclusion

This study has taken stock of the vast landscape of public web-based COVID-19 dashboards; this is a testament to the advancements in health information systems and digitalization of our societies, coupled with the responsibility and imperative to publicly report health information. As could be expected, the 158 dashboards in our sample, spanning a total of 53 countries, are diverse. They have different contexts and levels of focus, purposes, and audiences. They draw from various data sources, offer different content and use a range of ways—albeit at times limited—to break down data and to visualize, simplify, and interact with information. Their actionability also differs, signaling that their fitness for use by decision makers is not a guarantee. The number of dashboards appraised as highly actionable in the period of July 2020 when the dashboards in this study were assessed signals that work is still needed to optimize the use of dashboards. There is no one-size-fits-all template or model to accomplish this. Dashboards must be purpose-driven and context-specific. We urge those working on COVID-19 dashboards to consider the seven features identified in our study and adopt them as called for. By doing so, they stand to fully leverage the potential advantages of public reporting and its use for decision-making and behavior change needed to address the current pandemic.

Acknowledgments

We acknowledge the contributions of and are grateful to the following: Femke Jansen (Management Center Innsbruck); Omid Fekri (WHO Regional Office for Europe); Mekha Mathew (University of Oxford); and respondents to the initial dashboard sampling survey. This work was carried out by the Marie Skłodowska-Curie Innovative Training Network (HealthPros—Healthcare Performance Intelligence Professionals), which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement Nr. 765141.

Authors' Contributions

DI and EB contributed equally as first authors. DI, EB, SA, NK, and DK were involved in the design of the study. DI and EB coordinated the sampling of the dashboards, defined the inclusion and exclusion criteria, and performed the initial screening. DI and EB developed the dashboard assessment tool, which was piloted and validated by all authors. The panel of scorers (DI, EB, VB, OBF, KJG, TS, PK, NL, SL, BMT, JM, MP, AR, SW, CW, YY, and ZY) participated in data extraction, two rounds of scoring, and two consensus workshops. DI and EB coordinated the data extraction and organized and moderated the consensus workshops. DI and EB cleaned and analyzed the data and drafted the manuscript, with the supervision of SA, NK, and DK. All authors contributed to the interpretation of the results, provided critical comments, and edited, revised, and approved the manuscript in its final form for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Dashboard sampling survey.
[DOCX File , 18 KB - jmir_v23i2e25682_app1.docx]

Multimedia Appendix 2
Assessment tool.
[DOCX File , 48 KB - jmir_v23i2e25682_app2.docx]

Multimedia Appendix 3
Public web-based COVID-19 dashboards assessed.
[DOCX File , 65 KB - jmir_v23i2e25682_app3.docx]

Multimedia Appendix 4
Illustrative indicator titles by theme.

[DOCX File , 26 KB - jmir_v23i2e25682_app4.docx]

Multimedia Appendix 5
Summary of dashboard scoring.
[DOCX File , 33 KB - jmir_v23i2e25682_app5.docx]

References

1. Marshall MN, Romano PS, Davies HTO. How do we maximize the impact of the public reporting of quality of care? Int J Qual Health Care 2004 Apr 01;16 Suppl 1(suppl_1):i57-i63. [doi: 10.1093/intqhc/mzh013] [Medline: 15059988]
2. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
3. Strengthening and adjusting public health measures throughout the COVID-19 transition phases. World Health Organization. Copenhagen: WHO Regional Office for Europe; 2020 Apr 24. URL: https://apps.who.int/iris/bitstream/handle/10665/332467/WHO-EURO-2020-690-40425-54211-eng.pdf?sequence=1&isAllowed=y [accessed 2021-02-17]
4. Frieden TR. Government's role in protecting health and safety. N Engl J Med 2013 May 16;368(20):1857-1859. [doi: 10.1056/NEJMp1303819] [Medline: 23593978]
5. Smith P, Mossialos E, Papanicolas I, Leatherman S. Part 1: Principles of performance measurement. In: Smith P, Mossialos E, Papanicolas I, Leatherman S, editors. Performance Measurement for Health System Improvement Experiences, Challenges and Prospects. Cambridge, UK: Cambridge University Press; 2010:1-2.
6. Shekelle P. Public performance reporting on quality information. In: Smith PC, Mossialos E, Papanicolas I, Leatherman S, editors. Performance Measurement for Health System Improvement Experiences, Challenges and Prospects. Cambridge, UK: Cambridge University Press; 2010:537-551.
7. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. Med Care 2003 Jan;41(1 Suppl):I30-I38. [doi: 10.1097/00005650-200301001-00004] [Medline: 12544814]
8. Flattening the COVID-19 peak: containment and mitigation policies. OECD. Paris: OECD URL: https://read.oecd-ilibrary.org/view/?ref=124_124999-yt5ggxirhc&title=Flattening_the_COVID-19_peak-Containment_and_mitigation_policies [accessed 2021-02-17]
9. WHO checklist for influenza pandemic preparedness planning. World Health Organization. 2005. URL: https://www.who.int/influenza/resources/documents/FluCheck6web.pdf?ua=1 [accessed 2021-02-17]
10. Managing epidemics: key facts about major deadly diseases. World Health Organization. 2018. URL: https://apps.who.int/iris/handle/10665/272442 [accessed 2021-02-17]
11. Strengthening the health system response to COVID-19: recommendations for the WHO European Region. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/333072/WHO-EURO-2020-806-40541-54465-eng.pdf?sequence=1&isAllowed=y [accessed 2021-02-17]
12. Infodemic management. World Health Organization. URL: https://www.who.int/teams/risk-communication/infodemic-management [accessed 2021-02-17]
13. Few S. Dashboard confusion. Perceptual Edge. 2004 Mar 20. URL: https://www.perceptualedge.com/articles/ie/dashboard_confusion.pdf [accessed 2021-02-17]
14. Kennedy H. Simple data visualizations have become key to communicating about the COVID-19 pandemic, but we know little about their impact. London School of Economics and Political Sciences Impact Blog. 2020 May 04. URL: https://blogs.lse.ac.uk/impactofsocialsciences/2020/05/04/simple-data-visualisations-have-become-key-to-communicating-about-the-covid-19-pandemic-but-we-know-little-about-their-impact/ [accessed 2021-02-17]
15. Ghazisaeidi M, Safdari R, Torabi M, Mirzaee M, Farzi J, Goodini A. Development of performance dashboards in healthcare sector: key practical issues. Acta Inform Med 2015 Oct;23(5):317-321 [FREE Full text] [doi: 10.5455/aim.2015.23.317-321] [Medline: 26635442]
16. Damman OC, van den Hengel YK, van Loon AJM, Rademakers J. An international comparison of web-based reporting about health care quality: content analysis. J Med Internet Res 2010 Apr 13;12(2):e8 [FREE Full text] [doi: 10.2196/jmir.1191] [Medline: 20439252]
17. Weggelaar-Jansen AMJWM, Broekharst DSE, de Bruijne M. Developing a hospital-wide quality and safety dashboard: a qualitative research study. BMJ Qual Saf 2018 Dec 27;27(12):1000-1007 [FREE Full text] [doi: 10.1136/bmjqs-2018-007784] [Medline: 29950323]
18. Lee K, Jung SY, Hwang H, Yoo S, Baek HY, Baek R, et al. A novel concept for integrating and delivering health information using a comprehensive digital dashboard: An analysis of healthcare professionals' intention to adopt a new system and the trend of its real usage. Int J Med Inform 2017 Jan;97:98-108. [doi: 10.1016/j.ijmedinf.2016.10.001] [Medline: 27919400]
19. Kamadjeu R, Gathenji C. Designing and implementing an electronic dashboard for disease outbreaks response - Case study of the 2013-2014 Somalia Polio outbreak response dashboard. Pan Afr Med J 2017;27(Suppl 3):22 [FREE Full text] [doi: 10.11604/pamj.supp.2017.27.3.11062] [Medline: 29296157]

20. Meng Y, Zhang Y, Wang S, Wu S, Zhou H, Ke C, et al. Lessons learned in the development of a web-based surveillance reporting system and dashboard to monitor acute febrile illnesses in Guangdong and Yunnan Provinces, China, 2017-2019. Health Secur 2020 Jan;18(S1):S14-S22. [doi: 10.1089/hs.2019.0079] [Medline: 32004129]
21. Internet access. OECD. URL: https://data.oecd.org/ict/internet-access.htm [accessed 2021-02-17]
22. Nuti S, Noto G, Vola F, Vainieri M. Let's play the patients music: A new generation of performance measurement systems in healthcare. Manag Decis 2018 Oct 08;56(10):2252-2272. [doi: 10.1108/md-09-2017-0907]
23. Verheij R, Witvliet C, Jansen T, Hooiveld M, van Hilten H. Reuse of routine care data for policy and science: how things can be improved. Report in Dutch. Nivel. Utrecht: Nivel; 2019. URL: https://www.nivel.nl/sites/default/files/bestanden/Hergebruik_van_routine_zorgdata_voor_beleid_en_wetenschap.pdf [accessed 2021-02-21]
24. Galea S. The COVID conversation. Interference: International Review of Science (Special reports: SARS-CoV-2 and COVID-19). 2020 Jun 10. URL: https://inference-review.com/report/the-covid-conversation [accessed 2021-02-17]
25. WHO coronavirus disease (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-02-17]
26. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. John Hopkins Coronavirus Resource Centre. URL: https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 [accessed 2021-02-17]
27. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
28. COVID-19 economic recovery dashboard. Archived at archive.today. Deloitte. URL: http://archive.vn/qrJbw [accessed 2020-07-17]
29. Schiffmann A. nCoV2019 live. Archived at archive.today. URL: http://archive.vn/UWv29 [accessed 2020-07-21]
30. Health in the 21st century: putting data to work for stronger health systems. OECD. Paris: OECD; 2019. URL: https://read.oecd-ilibrary.org/social-issues-migration-health/health-in-the-21st-century_e3b23f8e-en#page1 [accessed 2021-02-17]
31. McDowell R. Signs to look for: criteria for developing and selecting fit for purpose indicators. Price Waterhouse Coopers. 2017. URL: https://ourlandandwater.nz/wp-content/uploads/2019/03/IWG_Indicators-thinkpiece-1.pdf [accessed 2021-02-17]
32. Smith P, Mossialos E, Papanicolas I. Performance measurement for health system improvement, experiences, challenges and prospects. World Health Organization Regional Office for Europe. 2008. URL: https://www.who.int/management/district/performance/PerformanceMeasurementHealthSystemImprovement2.pdf [accessed 2021-02-17]
33. Hibbard JH, Peters E. Supporting informed consumer health care decisions: data presentation approaches that facilitate the use of information in choice. Annu Rev Public Health 2003;24:413-433. [doi: 10.1146/annurev.publhealth.24.100901.141005] [Medline: 12428034]
34. Few S. What ordinary people need most from information visualization today. Visual Business Intelligence Newsletter. 2008 Aug. URL: http://www.perceptualedge.com/articles/visual_business_intelligence/what_people_need_from_infovis.pdf [accessed 2021-02-17]
35. Sadique MZ, Devlin N, Edmunds WJ, Parkin D. The effect of perceived risks on the demand for vaccination: results from a discrete choice experiment. PLoS One 2013;8(2):e54149 [FREE Full text] [doi: 10.1371/journal.pone.0054149] [Medline: 23408936]
36. Kringos D, Carinci F, Barbazza E, Bos V, Gilmore K, Groene O, HealthPros Network. Managing COVID-19 within and across health systems: why we need performance intelligence to coordinate a global response. Health Res Policy Syst 2020 Jul 14;18(1):80 [FREE Full text] [doi: 10.1186/s12961-020-00593-x] [Medline: 32664985]
37. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ 1995 Aug 05;311(7001):376-380 [FREE Full text] [doi: 10.1136/bmj.311.7001.376] [Medline: 7640549]
38. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. Am J Public Health 1984 Sep;74(9):979-983. [doi: 10.2105/ajph.74.9.979] [Medline: 6380323]
39. Van de Ven AH, Delbecq AL. The nominal group as a research instrument for exploratory health studies. Am J Public Health 1972 Mar;62(3):337-342. [doi: 10.2105/ajph.62.3.337] [Medline: 5011164]
40. International training network for healthcare performance intelligence professionals. HealthPros. URL: https://www.healthpros-h2020.eu/ [accessed 2021-02-17]
41. Lasswell HD. The structure and function of communication in society. In: Bryson L, editor. The Communication of Ideas. New York, NY: Harper and Row; 1948:136-139.
42. Kringos DS, Groene O, Johnsen SP. Training the first generation of health care performance intelligence professionals in Europe and Canada. Acad Med 2019 Jun;94(6):747-748. [doi: 10.1097/ACM.0000000000002694] [Medline: 31136339]
43. Charnock D. The DISCERN Handbook: Quality criteria for consumer health information on treatment choice. Oxford, UK: Radcliffe Medical Press; 1998.
44. Doak C, Doak L, Root J. SAM: Suitability Assessment of Materials for evaluation of health-related information for adults. Philadelphia, PA: J. B. Lippincott Company; 1996. URL: http://aspiruslibrary.org/literacy/SAM.pdf [accessed 2021-02-17]
45. Blanchet K, Nam SL, Ramalingam B, Pozo-Martin F. Governance and capacity to manage resilience of health systems: towards a new conceptual framework. Int J Health Policy Manag 2017 Aug 01;6(8):431-435. [doi: 10.15171/ijhpm.2017.36] [Medline: 28812842]

46. Sandmeyer B, Fraser I. New evidence on what works in effective public reporting. Health Serv Res 2016 Jun;51 Suppl 2:1159-1166 [FREE Full text] [doi: 10.1111/1475-6773.12502] [Medline: 27120996]
47. Peters E, Dieckmann N, Dixon A, Hibbard JH, Mertz CK. Less is more in presenting quality information to consumers. Med Care Res Rev 2007 Apr;64(2):169-190. [doi: 10.1177/1077558707060020301] [Medline: 17406019]
48. Raleigh VS, Foot C. Getting the measure of quality: opportunities and challenges. London: The King's Fund; 2010. URL: https://www.kingsfund.org.uk/sites/default/files/Getting-the-measure-of-quality-Veena-Raleigh-Catherine-Foot-The-Kings-Fund-January-2010.pdf [accessed 2021-02-17]
49. van den Berg MJ, Kringos DS, Marks LK, Klazinga NS. The Dutch Health Care Performance Report: seven years of health care performance assessment in the Netherlands. Health Res Policy Syst 2014 Jan 09;12:1 [FREE Full text] [doi: 10.1186/1478-4505-12-1] [Medline: 24405849]
50. Kreuter MW, Holt CL. How do people process health information? Applications in an age of individualized communication. Curr Dir Psychol Sci 2016 Jun 23;10(6):206-209. [doi: 10.1111/1467-8721.00150]
51. Kurtzman ET, Greene J. Effective presentation of health care performance information for consumer decision making: a systematic review. Patient Educ Couns 2016 Jan;99(1):36-43. [doi: 10.1016/j.pec.2015.07.030] [Medline: 26277826]
52. Ellis J. Six modest proposals. English Educ 2018 Mar;2(3):43-47. [doi: 10.1111/j.1754-8845.1968.tb00982.x]
53. Better information for improved health: a vision for health system use of data in Canada. Canadian Institute for Health Information. 2013. URL: https://www.infoway-inforoute.ca/en/component/edocman/1688-a-vision-for-health-system-use-of-data-in-canada-full-report/view-document?Itemid=0 [accessed 2021-02-17]
54. Hibbard JH, Greene J, Daniel D. What is quality anyway? Performance reports that clearly communicate to consumers the meaning of quality of care. Med Care Res Rev 2010 Jun;67(3):275-293. [doi: 10.1177/1077558709356300] [Medline: 20093399]
55. Ballard A. Promoting performance information use through data visualization: evidence from an experiment. Public Perform Manag Rev 2019 Apr 19;43(1):109-128. [doi: 10.1080/15309576.2019.1592763]
56. Webpage capture. archive.today. URL: http://archive.vn/ [accessed 2021-02-17]
57. King N. Using templates in the thematic analysis of text. In: Essential Guide to Qualitative Methods in Organizational Research. Newbury Park, CA: Sage Publications; 2004:256.
58. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
59. Alphabetical List of WHO Member States. World Health Organization. URL: https://www.who.int/choice/demography/by_country/en/ [accessed 2021-02-17]
60. #HowsMyFlattening. Archived at archive.today. URL: http://archive.vn/7T6MT [accessed 2020-07-14]
61. Covid Act Now: America's COVID Warning System. Archived at archive.today. URL: http://archive.vn/7hgsL [accessed 2020-07-14]
62. COVID-19 Sledilnik. Archived at archive.today. URL: http://archive.vn/szM4d [accessed 2020-07-09]
63. Global COVID-19 tracker and interactive charts. 1Point3Acres. Archived at archive.today. URL: http://archive.today/2020.07.14-101318/https://coronavirus.1point3acres.com/en [accessed 2020-07-14]
64. COVID-19 information for Switzerland. Archived at archive.today. URL: http://archive.vn/LmvZG [accessed 2020-07-08]
65. COVID-19 in Canada. Archived at archive.today. URL: http://archive.vn/m2dLJ [accessed 2020-07-13]
66. Information about COVID-19 precisely and accurately. Webpage in Indonesian. KawalCOVID19. URL: https://kawalcovid19.id/ [accessed 2020-07-17]
67. COVID Live. Archived at archive.today. URL: http://archive.vn/VtvRw [accessed 2020-07-16]
68. Coronavirus (COVID-19) in Australia. covid19data. Archived at archive.today. URL: http://archive.vn/CVeo8 [accessed 2020-07-08]
69. CoronaTracker. Archived at archive.today. URL: http://archive.vn/STFQO [accessed 2020-07-08]
70. Ustavimokorono. Archived at archive.today. URL: https://archive.vn/RZlgE [accessed 2020-07-09]
71. Koroonakaart. Archived at archive.today. URL: http://archive.vn/XxB8q [accessed 2020-07-17]
72. An interactive visualization of the exponential spread of COVID-19. 91-Divoc. Archived at archive.today. URL: http://archive.vn/fJ9tk [accessed 2020-07-17]
73. Little N. COVID-19 Tracker Canada. Archived at archive.today. URL: http://archive.vn/Lqk6L [accessed 2020-07-18]
74. COVID-19 Dashboard. CoronaBoard. Archived at archive.today. URL: http://archive.vn/aMu3a [accessed 2020-07-17]
75. COVID-19 data portal. Stats NZ. Archived at archive.today. URL: http://archive.vn/8Llz2 [accessed 2020-07-17]
76. COVID-19. Santé Publique France. Archived at archive.today. URL: http://archive.vn/nhsHy [accessed 2020-07-16]
77. Daily COVID-19 Dashboard. Ottawa Public Health. Archived at archive.today. URL: http://archive.vn/3ZL3k [accessed 2020-07-16]
78. COVID-19 dashboard. City of Vancouver. Archived at archive.today. URL: http://archive.vn/jPIlp [accessed 2020-07-20]
79. California SG. Tracking COVID-19 in California. Government of California. Archived at archive.today. URL: http://archive.vn/u41FC [accessed 2020-07-08]
80. Overvågning af COVID-19. Statens Serum Institut. URL: http://archive.vn/zzVm9 [accessed 2020-07-17]

81. COVID-19 in Spain. National Center of Epidemiology. Archvied at archive.today. URL: http://archive.today/2020.07.17-123630/https://cnecovid.isciii.es/covid19/ [accessed 2020-07-17]
82. Tracking the coronavirus. CBC News. Archived at archive.today. URL: http://archive.vn/p7Ix2 [accessed 2020-07-13]
83. Coronavirus world map: tracking the global outbreak. New York Times. URL: https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html [accessed 2021-08-01]
84. Ting I, Scott N, Workman M. Charting the COVID-19 spread in Australia. ABC News. Archived at archive.today. URL: http://archive.vn/o15kK [accessed 2020-07-17]
85. COVID-19: status of cases in Toronto. City of Toronto. Archived at archive.today. URL: http://archive.vn/NoyJh [accessed 2020-07-04]
86. Coronavirus (COVID-19) in the UK. Government of the United Kingdom. Archived at archive.today. URL: http://archive.vn/q5ilb [accessed 2020-07-17]
87. Coronavirus-Monitor. Berliner Morningpost. Archived at archive.today. URL: http://archive.vn/JF9dW [accessed 2020-07-16]
88. Colorado COVID-19 data. Colorado Department of Public Health and Environment and Colorado State Emergency Operations Center. Archived at archive.today. URL: http://archive.today/I3yam [accessed 2020-07-14]
89. LA County COVID-19 Surveillance Dashboard. Los Angeles County Department of Public Health. Archived at archive.today. URL: http://archive.vn/QBXTE [accessed 2020-07-17]
90. Coronavirus disease 2019 (COVID-19) outbreak. Maryland Department of Health. Archived at archive.today. URL: http://archive.vn/up0WC [accessed 2020-07-17]
91. Florida's COVID-19 Data and Surveillance Dashboard. Florida Department of Health. Archived at archive.today. URL: http://archive.vn/ZzNtp [accessed 2020-07-10]
92. COVID-19 data dashboard. Washington State Department of Health. Archived at archive.today. URL: http://archive.vn/94G5T [accessed 2020-07-13]
93. Papanicolas I, Smith P. Introduction. In: Health System Performance Comparison: An Agenda for Policy, Information and Research. Berkshire, UK: Open University Press; 2013.
94. Klazinga N, Stronks K, Delnoij D, Verhoeff A. Indicators without a cause. Reflections on the development and use of indicators in health care from a public health perspective. Int J Qual Health Care 2001 Dec;13(6):433-438. [doi: 10.1093/intqhc/13.6.433] [Medline: 11769744]
95. Hilarion P, Suñol R, Groene O, Vallejo P, Herrera E, Saura RM. Making performance indicators work: the experience of using consensus indicators for external assessment of health and social services at regional level in Spain. Health Policy 2009 Apr;90(1):94-103. [doi: 10.1016/j.healthpol.2008.08.002] [Medline: 18829129]
96. Compton-Phillips AT. The "Give a Darn" Method for Outcomes Measurement. NEJM Catalyst 2018 Feb 08 [FREE Full text]
97. Brehaut JC, Colquhoun HL, Eva KW, Carroll K, Sales A, Michie S, et al. Practice Feedback Interventions: 15 Suggestions for Optimizing Effectiveness. Ann Intern Med 2016 Mar 15;164(6):435-441. [doi: 10.7326/M15-2248] [Medline: 26903136]
98. Ploem C, Suurmond J. Registering ethnicity for covid-19 research: is the law an obstacle? BMJ 2020 Sep 01;370:m3367. [doi: 10.1136/bmj.m3367] [Medline: 32873550]
99. Rowley J. The wisdom hierarchy: representations of the DIKW hierarchy. J Inf Sci 2007 Feb 15;33(2):163-180. [doi: 10.1177/0165551506070706]

Abbreviations

NGT: nominal group technique
WHO: World Health Organization

Edited by R Kukafka; submitted 13.11.20; peer-reviewed by A Sales; comments to author 04.12.20; revised version received 09.12.20; accepted 31.01.21; published 24.02.21.

Please cite as:
Ivanković D, Barbazza E, Bos V, Brito Fernandes Ó, Jamieson Gilmore K, Jansen T, Kara P, Larrain N, Lu S, Meza-Torres B, Mulyanto J, Poldrugovac M, Rotar A, Wang S, Willmington C, Yang Y, Yelgezekova Z, Allin S, Klazinga N, Kringos D
Features Constituting Actionable COVID-19 Dashboards: Descriptive Assessment and Expert Appraisal of 158 Public Web-Based COVID-19 Dashboards
J Med Internet Res 2021;23(2):e25682
URL: https://www.jmir.org/2021/2/e25682
doi:10.2196/25682
PMID:33577467

Original Paper

Telemedicine Awareness, Knowledge, Attitude, and Skills of Health Care Workers in a Low-Resource Country During the COVID-19 Pandemic: Cross-sectional Study

Muhammed Elhadi[1], MBBCh; Ahmed Elhadi[1], MBBCh; Ahmad Bouhuwaish[2], MBBCh; Fatimah Bin Alshiteewi[1], MBBCh; Amna Elmabrouk[1], MBBCh; Ali Alsuyihili[1], MBBCh; Ayiman Alhashimi[3], MBBCh; Samer Khel[1], MBBCh; Alsafa Elgherwi[1], MBBCh; Ahmed Alsoufi[1], MBBCh; Ahmed Albakoush[1], MBBCh; Abdulmuez Abdulmalik[4], MBBCh

[1]Faculty of Medicine, University of Tripoli, Tripoli, Libyan Arab Jamahiriya

[2]Faculty of Medicine, Tobruk University, Tobruk, Libyan Arab Jamahiriya

[3]Faculty of Medicine, Al-Jabal Al Gharbi University, Gherian, Libyan Arab Jamahiriya

[4]Faculty of Medicine, Libyan International Medical University, Benghazi, Libyan Arab Jamahiriya

Corresponding Author:
Muhammed Elhadi, MBBCh
Faculty of Medicine
University of Tripoli
University Road, Furnaj
Tripoli, 13275
Libyan Arab Jamahiriya
Phone: 218 945196407
Email: muhammed.elhadi.uot@gmail.com

Abstract

Background: Since the onset of the COVID-19 pandemic, several health care programs intended to provide telemedicine services have been introduced in Libya. Many physicians have used these services to provide care and advice to their patients remotely.

Objective: This study aimed to provide an overview of physicians' awareness, knowledge, attitude, and skill in using telehealth services in Libya.

Methods: In this cross-sectional study, we administered a web-based survey to health care workers in Libya in May 2020. The questionnaire collected information on physicians' general demographic characteristics, ability to use a computer, and telemedicine awareness, knowledge, attitude, and skills.

Results: Among 673 health care workers who responded to the survey, 377 (56%) and 248 (36.8%) reported high awareness and high computer skill scores, respectively, for telemedicine. Furthermore, 582 (86.5%) and 566 (82.6%) health care workers reported high knowledge and high attitude scores, respectively. We observed no significant differences in awareness, knowledge, attitude, and skill scores among physicians employed at public, private, or both types of hospitals. We observed significant differences in the mean awareness (P<.001), attitude (P=.001), and computer skill scores (P<.001) , where the score distribution of the groups based on the ability to use computers was not similar. Knowledge scores did not significantly differ among the three groups (P=.37). Respondents with professional computer skills had significantly higher awareness (χ^2_3=14.5; P<.001) and attitude (χ^2_3=13.5; P=.001) scores than those without professional computer skills. We observed significant differences in the mean computer skill scores of the groups (χ^2_3=199.6; P<.001).

Conclusions: The consequences of the COVID-19 pandemic are expected to persist for a long time. Hence, policy programs such as telemedicine services, which aim to address the obstacles to medical treatment owing to physical distancing measures, will likely continue for a long time. Therefore, there is a need to train and support health care workers and initiate government programs that provide adequate and supportive health care services to patients in transitional countries.

(J Med Internet Res 2021;23(2):e20812) doi:10.2196/20812

KEYWORDS

attitude; awareness; coronavirus; COVID-19; knowledge; pandemic; skills; telemedicine

Introduction

SARS-CoV-2 was identified as the cause of severe viral pneumonia in Wuhan, Hubei Province, China, in December 2019. The World Health Organization declared COVID-19, the disease caused by this virus, a pandemic in March 2020 [1-4]. As of July 22, 2020, more than 15,000,000 COVID-19 cases and more than 617,000 COVID-19–related deaths have been recorded in 213 countries and territories [5].

Preventive measures have been implemented to reduce the potential exposure of individuals to the virus and to decrease the burden of COVID-19 transmission. These measures include hand washing, avoiding touching the face and eyes, cleaning and disinfecting surfaces and objects, wearing protective equipment including masks, and most importantly, social distancing by staying at home, avoiding close physical contact with others, and maintaining adequate interpersonal distance [6-8].

The COVID-19 pandemic has affected traditional in-person health care services. New strategies and approaches, including telemedicine and telehealth services, must be implemented to facilitate video consultations and the use of mobile apps, with the goals of delivering medical advice, diagnosis, and treatment to reduce the risk of infection [9,10]. Telemedicine and telehealth services are especially important for individuals without COVID-19 seeking medical treatment during the pandemic, especially those with chronic or acute illnesses. Several countries had begun increasing the use of telemedicine services even prior to the COVID-19 pandemic [11-16]. Telemedicine was effective during previous outbreaks including those of the Ebola virus, Zika virus, and the severe acute respiratory syndrome coronavirus [17,18].

Telemedicine is defined as the use of web-based resources and electronic information along with advanced digital network technology to promote long-distance professional health services, disseminate medical safety reports, deliver health-related education to the public, and carry out public health monitoring [19,20].

Since the onset of the COVID-19 pandemic, several health care service programs aiming to provide telemedicine services have been introduced in Libya. Even with their recent introduction, many physicians have used these services to provide care and advice to patients remotely. There is a need to assess the awareness, knowledge, attitude, and skills of physicians who use telemedicine services. The skilled use of telemedicine by physicians benefits patients, who can receive a wide range of health care services without the need to visit the clinic. Telemedicine services can reduce the risk of infection among populations by facilitating social distancing and decreasing the burden on the hospitals that provide care to individuals with COVID-19 [21-23].

This study aimed to provide an overview of physicians' awareness, knowledge, attitude, and skills of telemedicine—a technology that has been recently introduced in transitional countries. Furthermore, we investigated the effects of COVID-19 on this technology to determine whether it is adequately applicable during the ongoing COVID-19 pandemic.

Methods

Study Design

In this cross-sectional study, we administered a web-based survey to health care workers in Libya in May 2020 via text messages and email. Responses were anonymous without any identifying data. This study focused on health care workers who were officially registered with Libyan hospitals during the COVID-19 pandemic and were involved in patient care. The first section of the survey collected baseline demographic information and characteristics of the health care workers, including age, gender, years of experience, employment status, department of work, and computer skills, and there were several general questions related to telemedicine experience. The second section contained a survey instrument that was previously developed by Zayapragassarazan et al [24], which contains four detailed subsections.

Awareness Subsection

This subsection consisted of 12 items on telemedicine, which were used to assess awareness. Each item was assessed with a 3-point scale with scores of 0-2 (0=don't know, 1=heard of it, 2=know about it). The overall score could range from a minimum of 0 to a maximum of 24.

Knowledge Subsection

This subsection consisted of 11 items that assessed the physician's level of telemedicine knowledge. Each item was assessed with a 2-point scale with scores of 0 and 1 (0=yes, 1=no). The overall score could range from a minimum of 0 to a maximum of 11.

Attitude Subsection

This subsection consisted of 11 items used to determine the attitude of the respondents toward telemedicine. Each item was assessed with a 5-point Likert scale with scores ranging 0-4 (0=strongly disagree, 1=disagree, 2=undecided, 3=agree, 4=strongly agree). The overall score could range from a minimum of 0 to a maximum of 44.

Computer Skills Subsection

This subsection consisted of 13 items used to determine the physician's level of information technology and computer skills. Each item was assessed with a 4-point scale with scores of 0-3 (0=unskilled, 1=learner, 2=mediocre, 3=expert). The overall score could range from a minimum of 0 to a maximum of 39.

Assessment of the Effect of COVID-19 on Telemedicine

The third section included several questions on the effects of the spread of COVID-19 on telemedicine, the role of telemedicine in reducing in-person consultations, the time spent on practicing telemedicine, the usability of telemedicine based

on local internet settings, and the role of telemedicine in helping physicians avoid or decrease the risk of infection potentially resulting from internal conflicts in Libya.

Raw scores of each subsection of the awareness-knowledge-attitude-skill questionnaire were converted to percentages and compared with other group variables. We considered scores of ≤49% as low, 50%-70% as average, and ≥71% as high. The survey was conducted in accordance with the Checklist for Reporting Results of Internet E-Surveys [25].

Statistical Analysis

Descriptive statistics related to the general demographic characteristics of the physicians were calculated. Frequency and percentage values were used to describe these variables. Mean (SD) values were used to describe continuous data. None of the major continuous outcomes, including survey scores, were normally distributed. Therefore, we analyzed them using the Mann-Whitney *U* test to determine whether there were significant differences between the two groups and the score of each section. We performed the Kruskal-Wallis *H* test to compare groups with ≥3 independent variables. The chi-square test was used to determine potential associations between the categorical groups. Statistical analysis was performed using SPSS software (version 25.0, IBM Corp). A *P* value of ≤.05 was considered significant.

Ethical Consideration

The study procedures comply with the ethical standards of the relevant national and institutional committee on human experimentation, with the Helsinki Declaration of 1975, as revised in 2008. Ethical approval (approval# 109.3-2020) for this study was obtained from the Bioethics Committee at the Biotechnology Research Centre (Tripoli, Libya). The survey was conducted anonymously, and all participants provided written informed consent before participating in the study.

Results

Telemedicine Awareness, Knowledge, Attitude, and Skills of Health Care Workers

From among 800 health care workers who received the survey, 673 responded (response rate=84.1%). Among them, 299 (44.4%) were male and 374 (55.6%) were female. The participant cohort comprised 288 (42.8%) specialists or senior physicians and 385 (57.2%) physicians in training. Table 1 summarizes the baseline demographic characteristics of the study participants.

Most participants (n=631, 93.8%) had an average or high level of computer and internet skills. Most participants (n=525, 78%) had participated in training programs on telemedicine during the COVID-19 pandemic. Most participants (n=639, 94.9%) expressed their willingness to participate in such training courses.

The questionnaires assessed the physicians' awareness, knowledge, attitude, and skills of telemedicine. The assessment scales had a high level of internal consistency (Cronbach α=.823 for awareness, Cronbach α=.735 for knowledge, Cronbach α=.910 for attitude, and Cronbach α=.950 for skills).

Interestingly, 399 (59.3%) participants perceived telemedicine as a helpful tool for elderly patients, and 486 (72.2%) speculated that telemedicine could be used by physicians to obtain patients' electrocardiography and x-ray reports for medical consultation. Some participants (n=178, 26.4%) speculated that telemedicine technologies can facilitate direct consultations. Only 28 (4.2%) participants speculated that laboratory findings could be shared with patients through telemedicine technologies, and only 81 (12%) participants speculated that follow-up and treatment response evaluation could be performed using telemedicine technology.

Many participants (n=580, 86.2%) did not agree that patients should have access to information regarding their medical condition. However, 573 (85.1%) participants believed that telemedicine could help bridge the gap between primary and secondary care by fostering communication among physicians, consultants, and nurses. Interestingly, 254 (37.7%) participants had little or no experience using emails and sending files.

Data on physicians' awareness, knowledge, attitude, and skills of telemedicine are summarized in Table 2. In total, 377 (56%) participants had high awareness scores, 582 (86.5%) had high knowledge scores, 556 (82.6%) had high attitude scores, and 248 (36.8%) had high scores for computer skills of telemedicine technology for all specialties. Overall, 132 (19.6%) participants were not familiar with telemedicine technology, and 28 (4.2%) had low telemedicine knowledge. In addition, 16 (2.4%) participants had low attitude scores.

Table 1. Baseline demographic characteristics of the study participants (N=673).

Variables	Total	Females (n=374)	Males (n=299)	*P* value
Age range (years), n (%)				.007[a]
<30	109 (16.2)	48 (12.8)	61 (20.4)	
30-40	442 (65.7)	247 (66.0)	195 (65.2)	
>40	122 (18.1)	79 (21.1)	43 (14.4)	
Experience (years), mean (SD)	8.78 (8.09)	9.6 (8.54)	7.76 (7.38)	.001[a]
Employment status, n (%)				<.001[b]
Public sector	335 (49.8)	150 (40.1)	185 (61.9)	
Private sector	82 (12.2)	49 (13.1)	33 (11.0)	
Both	256 (38.0)	175 (46.8)	81 (27.1)	
Ability to use computers, n (%)				.004[a]
Beginner	42 (6.2)	15 (4.0)	27 (9.0)	
Average	452 (67.2)	246 (65.8)	206 (68.9)	
Professional	179 (26.6)	113 (30.2)	66 (22.1)	
Received training for telemedicine system, n (%)				.37
Yes	525 (78.0)	287 (76.7)	238 (79.6)	
No	148 (22.0)	87 (23.3)	61 (20.4)	
Availability of a telemedicine unit in your department, n (%)				.61
Yes	525 (78.0)	289 (77.3)	236 (78.9)	
No	148 (22.0)	85 (22.7)	63 (21.1)	
Department, n (%)				<.001[b]
Internal medicine (including all subspecialties)	179 (26.6)	101 (27.0)	78 (26.1)	
Pediatrics	74 (11.0)	33 (8.8)	41 (13.7)	
Surgery (including all subspecialties)	160 (23.8)	131 (35.0)	29 (9.7)	
Gynecology and obstetrics	72 (10.7)	3 (0.8)	69 (23.1)	
Dermatology	36 (5.3)	11 (2.9)	25 (8.4)	
Family medicine	64 (9.5)	38 (10.2)	26 (8.7)	
Psychiatry	25 (3.7)	14 (3.7)	11 (3.7)	
Other	63 (9.4)	43 (11.5)	20 (6.7)	

[a]*P*<.05.

[b]*P*<.001.

Table 2. Different levels of awareness, knowledge, attitude, and skills among health care workers in Libya (N=673).

Degree	Awareness	Knowledge	Attitude	Skills
Low (≤49%), n (%)	132 (19.6)	28 (4.2)	16 (2.4)	226 (33.6)
Average (50%-70%), n (%)	164 (24.4)	63 (9.4)	101 (15.0)	199 (29.6)
High (≥71%), n (%)	377 (56.0)	582 (86.5)	556 (82.6)	248 (36.8)

Table 3 compares awareness and knowledge scores, and Table 4 compares the attitude and skill scores among the study participants. The Mann-Whitney *U* test was performed to investigate significant differences in the awareness, knowledge, attitude, and computer skill scores between males and females. Females had significantly higher mean awareness (*P*=.02) and computer skill scores (*P*<.001) than males. However, we observed no significant differences in the mean knowledge (*P*=.55) and attitude (*P*=.99) scores between males and females.

Table 3. Differences in the awareness and knowledge scores of the study participants based on baseline demographic characteristics (N=673).

Variables	Total, n (%)	Awareness scores (range 0-100)			Knowledge scores (range 0-100)		
		Mean (SD)	*U*[a] or *H*[b] value	*P* value	Mean (SD)	*U* or *H* value	*P* value
Gender			49926.5[a]	.02[c]		54460.5[a]	.55
Male	109 (16.2)	69.9 (24.8)			83.0 (16.4)		
Female	442 (65.7)	66.5 (23.5)			83.5 (17.1)		
Employment status			2.79[b]	.25		2.16 [b]	.34
Public sector	335 (49.8)	68.0 (24.9)			83.1 (16.6)		
Private sector	82 (12.2)	73.1 (21.2)			86.0 (13.9)		
Both	256 (38.0)	67.5 (24.3)			82.6 (17.7)		
Ability to use the computer			41.97[b]	<.001[d]		1.98[b]	.37
Beginner	42 (6.2)	49.6 (28.1)			77.3 (24.9)		
Average	452 (67.2)	67.1 (23.8)			83.3 (16.2)		
Professional	179 (26.6)	76.3 (21.4)			84.6 (15.5)		
Department			8.99[b]	.25		10.68[b]	.15
Internal medicine	179 (26.6)	67.4 (25.9)			82.4 (17.3)		
Pediatrics	74 (11.0)	65.6 (24.6)			83.7 (15.3)		
Surgical specialties	160 (23.8)	66.9 (26.2)			80.8 (19.6)		
Gynecology and obstetrics	72 (10.7)	73.4 (18.2)			85.6 (15.8)		
Dermatology	36 (5.3)	70.4 (20.4)			85.6 (12.4)		
Family medicine	64 (9.5)	73.5 (20.8)			86.4 (15.2)		
Psychiatry	25 (3.7)	73.7 (24.5)			89.5 (8.2)		
Other	63 (9.4)	64.3 (24.5)			82.3 (15.2)		

[a]Mann-Whitney *U* test for pairwise comparison of independent variables.

[b]Kruskal-Wallis *H* test for multiple-group comparison of independent variables.

[c]*P*<.05.

[d]*P*<.001.

Table 4. Differences in the attitude and computer skill scores of the study participants based on their baseline demographic characteristics (N=673).

Variables	Total, n (%)	Attitude scores (range 0-100)			Computer skill scores (range 0-100)		
		Mean (SD)	*U*[a] or *H*[b] value	*P* value	Mean (SD)	*U* or *H* value	*P* value
Gender			55878.50[a]	.99		43475.50[a]	<.001
Male	109 (16.2)	79.8 (12.7)			56.4 (25.3)		
Female	442 (65.7)	79.8 (14.3)			66.1 (22.3)		
Employment status			1.93[b]	.38		4.17[b]	.12
Public sector	335 (49.8)	80.3 (13.5)			60.1 (24.8)		
Private sector	82 (12.2)	80.7 (13.4)			66.1 (20.2)		
Both	256 (38.0)	78.9 (13.7)			62.7 (24.3)		
Ability to use the computer			13.46[b]	.001[c]		199.62[b]	<.001[d]
Beginner	42 (6.2)	72.6 (15.7)			37.9 (21.6)		
Average	452 (67.2)	79.6 (13.3)			55.8 (20.5)		
Professional	179 (26.6)	81.9 (13.3)			82.4 (19.3)		
Department			16.67[b]	.02[c]		22.87[b]	.002[c]
Internal medicine	179 (26.6)	78.3 (13.9)			59.8 (25.8)		
Pediatrics	74 (11.0)	79.6 (13.6)			56.5 (25.6)		
Surgical specialties	160 (23.8)	78.9 (15.7)			67.1 (21.9)		
Gynecology and obstetrics	72 (10.7)	83.1 (10.5)			56.5 (23.8)		
Dermatology	36 (5.3)	80.4 (14.0)			55.3 (25.4)		
Family medicine	64 (9.5)	82.8 (13.2)			63.6 (22.7)		
Psychiatry	25 (3.7)	83.1 (8.9)			71.9 (24.7)		
Other	63 (9.4)	78.4 (10.9)			64.1 (23.1)		

[a]Mann-Whitney *U* test for pairwise comparison of independent variables.

[b]Kruskal-Wallis *H* test for multiple-group comparison of independent variables.

[c]*P*<.05.

[d]*P*<.001.

We performed the Kruskal-Wallis *H* test to investigate differences in the awareness, knowledge, attitude, and skill scores of other categorical groups, including those based on employment status, ability to use computers, and departments of work. We observed no significant differences in the awareness (*P*=.25), knowledge (*P*=.34), attitude (*P*=.38), and computer skill (*P*=.12) scores among health care workers employed at public, private, or both types of hospitals. We observed significant differences in the mean awareness (*P*<.001), attitude (*P*=.001), and computer skill (*P*<.001) scores, where the score distribution based on the ability to use computers was not similar among these groups. We observed no significant differences in the knowledge scores of the three groups based on employment status (*P*=.34). Participants with professional computer skills had significantly higher awareness (χ^2_3=14.5; *P*<.001) and attitude (χ^2_3=13.5; *P*=.001) scores than those without professional computer skills. Furthermore, we observed significant differences in the mean computer skill scores among different groups (χ^2_3=199.6; *P*<.001). Overall, those who were more capable of using computers displayed significantly higher computer skill scores than those who were lesser capable of using computers.

We observed significant differences in attitude scores (χ^2_8=16.7; *P*=.02) and computer skills (χ^2_8=22.9; *P*=.002) among health care workers in different departments. In particular, those working in the departments of gynecology and obstetrics, family medicine, and psychiatry had significantly higher attitude scores (*P*=.02), while those in the department of psychiatry and those in surgical specialties had significantly higher computer skill scores (*P*=.002) than those of their counterparts in other departments. However, awareness (*P*=.25) and knowledge (*P*=.15) scores did not significantly differ among health care workers in different departments.

Effects of COVID-19 on Telemedicine

Among the study participants, 638 (94.8%) thought that telemedicine technology could be used to limit the spread of COVID-19. Additionally, 630 (93.6%) participants thought that telemedicine could help reduce hospital visits to avoid COVID-19 transmission. However, only 283 (42.1%) participants thought that telemedicine could replace traditional

medical visits during the COVID-19 pandemic. In addition, 616 (91.5%) participants thought that the telemedicine system could facilitate patient communication and cooperation with physicians during the COVID-19 pandemic. Furthermore, 622 (92.4%) participants thought that the use of telemedicine and remote health care systems saved a lot of time that was otherwise lost in hospitals and clinics and that these remote services could more rapidly provide medical advice. However, only 437 (64.9%) participants thought that internet services in Libya met the demand for uninterrupted telemedicine services. Moreover, 121 (18%) participants thought that telemedicine could help physicians avoid the issues of conflict and civil war in Libya and decrease the risk of infection arising from these internal conflicts. Interestingly, 575 (85.4%) participants thought that telemedicine could potentially prevent physicians who worked in unsafe environments from sensing insecurity during the civil war. Additionally, 606 (90%) participants thought that telemedicine could help physicians avoid the risk of contracting and transmitting COVID-19. However, only 385 (57.2%) participants thought that the telemedicine service technologies would help improve the financial status of physicians.

Discussion

Principal Findings

Telemedicine has been introduced recently in transitional countries, including Libya, to provide remote health care services, especially to patients with chronic diseases. In this study, 377 (56%), 582 (86.5%), and 566 (82.6%) physicians responding to our survey had high awareness, knowledge, and attitude levels for telemedicine, respectively; however, only 248 (36.8%) participants had adequate or high computer skills.

The COVID-19 pandemic has placed greater emphasis on telemedicine technology as a means of providing adequate care to patients without increasing the risk of SARS-CoV-2 transmission to patients during in-hospital clinic visits. Additionally, telemedicine benefits health care workers because it decreases the risk of infection among them, reduces stress in the hospital, provides an adequate tool for treatment and follow-up evaluation of their patients, and provides a suitable approach for providing mental health services.

This study provides an overview of physicians' awareness, knowledge, attitude, and skill levels of telemedicine in a transitional country that is afflicted with a civil war, a financial crisis, and a devastating health care system during the COVID-19 pandemic [26-28].

Only 179 (26.6%) participants had professional computer skills, while 452 (67.2%) and 42 (6.2%) participants had average and beginner computer skill levels, respectively. Thus, overall, the computer and information technology skills of the physicians were inadequate. Such skills are crucial for the productive use of telemedicine services. Therefore, we recommend software and computer skill training programs for physicians who are newly introduced to telemedicine technology in transitional countries. Nonetheless, the approach may be difficult even with these skills. Among all our participants, 236 (35.1%) indicated that the internet services in Libya were not sufficient to meet patient needs through telemedicine technology owing to electricity and internet interruptions in Libya after the civil war [29]. This remains yet another challenge for telemedicine service providers, requiring immediate attention from internet providers and telephone companies [30]. Many patients have complained about the unavailability of these services owing to their high demand during the COVID-19 pandemic.

Most study participants (n=556, 82.6%) had higher attitude scores. This is important because attitude represents how telemedicine was perceived by health care workers [31]. The attitude of health care workers is an important factor in understanding and accepting telemedicine technologies. For such acceptance, the program developer needs to train health care workers and make the telemedicine programs usable for them [32].

Knowledge and attitude are important for the acceptance of telemedicine by health care workers. In this study, 582 (86.5%) participants had high knowledge scores; this value is higher than that reported previously in Germany (where approximately 63% of individuals had some telemedicine knowledge) [33]. Another study conducted in Lagos [34] reported that only 60.9% of participants had telemedicine knowledge. Furthermore, another study conducted in Pakistan [35] reported an average level of telemedicine knowledge among health care workers. Further initiatives and commitments are required to expand the use of telemedicine and increase its efficiency, especially among health care professionals in a postconflict setting. There is a need for greater understanding of telemedicine among health care practitioners, particularly those who are new telemedicine users. Hence, it is important to address the financial and physical access–related obstacles to telemedicine [36].

There are several constraints in the implementation of telemedicine in transitional countries, especially in emergency settings, notwithstanding the implementation of strategies to reduce the risk of exposure to the virus during the COVID-19 pandemic. Certain technical issues and the need to train health care workers to provide adequate telemedicine services need to be addressed. The COVID-19 pandemic has put a strain on hospital health care workers and has posed challenges to politicians, managers, and practitioners with regard to the limits of the broader health care infrastructure and assumptions that have restricted its potential for swift and innovative reforms [30,37,38].

Limitations

Our study was limited to Libya; hence, our findings may not be generalizable to the populations of other countries. Furthermore, the cross-sectional design of our study does not allow for drawing concrete conclusions establishing causation between variables. In addition, our study does not provide details regarding the types of telemedicine technologies and how they can be implemented, along with patients' attitude toward telemedicine. Further studies on telemedicine are required to provide qualitative measures of health care services to patients, to assess the access to health care services in other transitional countries, and to determine whether these telemedicine adaptations benefit patients, especially those with chronic illnesses and psychiatric morbidities [39-41]. Additionally, there

is a need to determine the cost benefit of these services and whether modifications are required in these services.

Conclusions

The consequences of the COVID-19 pandemic are expected to be long lasting. Hence, policy programs, such as telemedicine services, which aim to address obstacles to medical treatment owing to physical distancing measures, are also required to be provided over a long term [42]. Therefore, there is a need to train and support health care workers and initiate government programs that provide adequate and supportive health care services to patients in transitional countries and to respond to issues regarding the access to these services, including internet access and advertisements, and social programs that help patients understand how to use these services.

Acknowledgments

This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of Interest

None declared.

References

1. Chan JF, Yuan S, Kok K, To KK, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 2020 Feb 15;395(10223):514-523 [FREE Full text] [doi: 10.1016/S0140-6736(20)30154-9] [Medline: 31986261]
2. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
3. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020 Mar;579(7798):270-273 [FREE Full text] [doi: 10.1038/s41586-020-2012-7] [Medline: 32015507]
4. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed 2020 Mar 19;91(1):157-160 [FREE Full text] [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
5. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
6. Coronavirus disease (COVID-19) advice for the public. World Health Organization. 2021. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public [accessed 2021-02-22]
7. Yan Y, Shin WI, Pang YX, Meng Y, Lai J, You C, et al. The First 75 Days of Novel Coronavirus (SARS-CoV-2) Outbreak: Recent Advances, Prevention, and Treatment. Int J Environ Res Public Health 2020 Mar 30;17(7):2323 [FREE Full text] [doi: 10.3390/ijerph17072323] [Medline: 32235575]
8. Orozco-Beltran D, Sánchez-Molla M, Sanchez JJ, Mira JJ, ValCrònic Research Group. Telemedicine in Primary Care for Patients With Chronic Conditions: The ValCrònic Quasi-Experimental Study. J Med Internet Res 2017 Dec 15;19(12):e400. [doi: 10.2196/jmir.7677] [Medline: 29246881]
9. Ohannessian R, Duong TA, Odone A. Global Telemedicine Implementation and Integration Within Health Systems to Fight the COVID-19 Pandemic: A Call to Action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
10. Lin C, Tseng W, Wu J, Tay J, Cheng M, Ong H, et al. A Double Triage and Telemedicine Protocol to Optimize Infection Control in an Emergency Department in Taiwan During the COVID-19 Pandemic: Retrospective Feasibility Study. J Med Internet Res 2020 Jun 23;22(6):e20586 [FREE Full text] [doi: 10.2196/20586] [Medline: 32544072]
11. Liu S, Yang L, Zhang C, Xiang Y, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. Lancet Psychiatry 2020 Apr;7(4):e17-e18 [FREE Full text] [doi: 10.1016/S2215-0366(20)30077-8] [Medline: 32085841]
12. Li W, Yang Y, Liu Z, Zhao Y, Zhang Q, Zhang L, et al. Progression of Mental Health Services during the COVID-19 Outbreak in China. Int J Biol Sci 2020;16(10):1732-1738 [FREE Full text] [doi: 10.7150/ijbs.45120] [Medline: 32226291]
13. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai X, et al. The Role of Telehealth in Reducing the Mental Health Burden from COVID-19. Telemed J E Health 2020 Apr;26(4):377-379. [doi: 10.1089/tmj.2020.0068] [Medline: 32202977]
14. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
15. Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
16. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine During the COVID-19 Pandemic: Experiences From Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]

17. Tom-Aba D, Olaleye A, Olayinka AT, Nguku P, Waziri N, Adewuyi P, et al. Innovative Technological Approach to Ebola Virus Disease Outbreak Response in Nigeria Using the Open Data Kit and Form Hub Technology. PLoS One 2015;10(6):e0131000 [FREE Full text] [doi: 10.1371/journal.pone.0131000] [Medline: 26115402]
18. Chang T, Lee J, Wu S. The telemedicine and teleconsultation system application in clinical medicine. In: Conf Proc IEEE Eng Med Biol Soc. 2004 Presented at: The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 1-5 September 2004; San Francisco, CA p. 3392-3395. [doi: 10.1109/IEMBS.2004.1403953]
19. Wootton R. Recent advances: Telemedicine. BMJ 2001 Sep 08;323(7312):557-560 [FREE Full text] [doi: 10.1136/bmj.323.7312.557] [Medline: 11546704]
20. Ryu S. Telemedicine: Opportunities and Developments in Member States: Report on the Second Global Survey on eHealth 2009 (Global Observatory for eHealth Series, Volume 2). Healthc Inform Res 2012;18(2):153 [FREE Full text] [doi: 10.4258/hir.2012.18.2.153]
21. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
22. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, et al. Rapid response to COVID-19: health informatics support for outbreak management in an academic health system. J Am Med Inform Assoc 2020 Jun 01;27(6):853-859 [FREE Full text] [doi: 10.1093/jamia/ocaa037] [Medline: 32208481]
23. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth Interventions to Support Self-Management of Long-Term Conditions: A Systematic Metareview of Diabetes, Heart Failure, Asthma, Chronic Obstructive Pulmonary Disease, and Cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
24. Zayapragassarazan Z, Kumar S. Awareness, Knowledge, Attitude and Skills of Telemedicine among Health Professional Faculty Working in Teaching Hospitals. J Clin Diagn Res 2016 Mar;10(3):JC01-JC04 [FREE Full text] [doi: 10.7860/JCDR/2016/19080.7431] [Medline: 27134899]
25. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
26. Elhadi M, Msherghi A, Alkeelani M, Alsuyihili A, Khaled A, Buzreg A, et al. Concerns for low-resource countries, with under-prepared intensive care units, facing the COVID-19 pandemic. Infect Dis Health 2020 Nov;25(4):227-232 [FREE Full text] [doi: 10.1016/j.idh.2020.05.008] [Medline: 32631682]
27. Elhadi M, Msherghi A, Alkeelani M, Zorgani A, Zaid A, Alsuyihili A, et al. Assessment of Healthcare Workers' Levels of Preparedness and Awareness Regarding COVID-19 Infection in Low-Resource Settings. Am J Trop Med Hyg 2020 Aug;103(2):828-833 [FREE Full text] [doi: 10.4269/ajtmh.20-0330] [Medline: 32563273]
28. Elhadi M, Msherghi A, Elgzairi M, Alsuyihili A, Elkhafeefi F, Bouhuwaish A, et al. Assessment of the preparedness of obstetrics and gynecology healthcare systems during the COVID-19 pandemic in Libya. Int J Gynaecol Obstet 2020 Sep;150(3):406-408. [doi: 10.1002/ijgo.13273] [Medline: 32538491]
29. Elhadi M, Msherghi A. COVID-19 and civil war in Libya: the current situation. Pathog Glob Health 2020 Jul;114(5):230-231. [doi: 10.1080/15575330.2020.1769292] [Medline: 32432976]
30. Angood PB. Telemedicine, the Internet, and world wide web: overview, current status, and relevance to surgeons. World J Surg 2001 Nov;25(11):1449-1457. [doi: 10.1007/s00268-001-0130-4] [Medline: 11760749]
31. Ayatollahi H, Sarabi FZP, Langarizadeh M. Clinicians' Knowledge and Perception of Telemedicine Technology. Perspect Health Inf Manag 2015;12:1c [FREE Full text] [Medline: 26604872]
32. Ward R, Stevens C, Brentnall P, Briddon J. The attitudes of health care staff to information technology: a comprehensive review of the research literature. Health Info Libr J 2008 Jun;25(2):81-97 [FREE Full text] [doi: 10.1111/j.1471-1842.2008.00777.x] [Medline: 18494643]
33. Tomás C, Oliveira E, Sousa D, Uba-Chupel M, Furtado G, Rocha C, et al. Proceedings of the 3rd IPLeiria's International Health Congress. In: BMC Health Serv Res. 2016 Jul 06 Presented at: 3rd IPLeiria's International Health Congress; 6-7 May 2016; Leiria, Portugal p. 200 URL: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-016-1423-5 [doi: 10.1186/s12913-016-1423-5]
34. Joseph L, Shittu L, Adesanya O, Mathew C, Izegbu, Oluwole A. Knowledge and perception of health workers towards tele-medicine application in a new teaching hospital in Lagos. Sci Res Essays 2007;2(1):016-019 [FREE Full text]
35. Ashfaq A, Memon SF, Zehra A, Barry S, Jawed H, Akhtar M, et al. Knowledge and Attitude Regarding Telemedicine Among Doctors in Karachi. Cureus 2020 Feb 09;12(2):e6927 [FREE Full text] [doi: 10.7759/cureus.6927] [Medline: 32190480]
36. Woodward A, Fyfe M, Handuleh J, Patel P, Godman B, Leather A, et al. Diffusion of e-health innovations in 'post-conflict' settings: a qualitative study on the personal experiences of health workers. Hum Resour Health 2014 Apr 23;12:22 [FREE Full text] [doi: 10.1186/1478-4491-12-22] [Medline: 24754997]
37. Lonergan PE, Washington Iii SL, Branagan L, Gleason N, Pruthi RS, Carroll PR, et al. Rapid Utilization of Telehealth in a Comprehensive Cancer Center as a Response to COVID-19: Cross-Sectional Analysis. J Med Internet Res 2020 Jul 06;22(7):e19322 [FREE Full text] [doi: 10.2196/19322] [Medline: 32568721]

38. Xu H, Huang S, Qiu C, Liu S, Deng J, Jiao B, et al. Monitoring and Management of Home-Quarantined Patients With COVID-19 Using a WeChat-Based Telemedicine System: Retrospective Cohort Study. J Med Internet Res 2020 Jul 02;22(7):e19514. [doi: 10.2196/19514] [Medline: 32568727]
39. Ming W, Mackillop LH, Farmer AJ, Loerup L, Bartlett K, Levy JC, et al. Telemedicine Technologies for Diabetes in Pregnancy: A Systematic Review and Meta-Analysis. J Med Internet Res 2016 Nov 09;18(11):e290 [FREE Full text] [doi: 10.2196/jmir.6556] [Medline: 27829574]
40. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for Future Mobile-Health (mHealth): A Systematic Review of Mobile Phone and Web-Based Text Messaging in Mental Health. J Med Internet Res 2016 Jun 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
41. Jiang X, Ming W, You JH. The Cost-Effectiveness of Digital Health Interventions on the Management of Cardiovascular Diseases: Systematic Review. J Med Internet Res 2019 Jun 17;21(6):e13166 [FREE Full text] [doi: 10.2196/13166] [Medline: 31210136]
42. O'Brien M, McNicholas F. The Use of Telepsychiatry During COVID-19 and Beyond. Ir J Psychol Med 2020 May 21:1-17. [doi: 10.1017/ipm.2020.54]

Edited by G Eysenbach, G Fagherazzi; submitted 29.05.20; peer-reviewed by R Ohannessian, B Clay, H Ayatollahi; comments to author 13.07.20; revised version received 22.07.20; accepted 03.09.20; published 25.02.21.

Please cite as:
Elhadi M, Elhadi A, Bouhuwaish A, Bin Alshiteewi F, Elmabrouk A, Alsuyihili A, Alhashimi A, Khel S, Elgherwi A, Alsoufi A, Albakoush A, Abdulmalik A
Telemedicine Awareness, Knowledge, Attitude, and Skills of Health Care Workers in a Low-Resource Country During the COVID-19 Pandemic: Cross-sectional Study
J Med Internet Res 2021;23(2):e20812
URL: https://www.jmir.org/2021/2/e20812
doi:10.2196/20812
PMID:33600350

Original Paper

Perception of COVID-19 Physical Distancing Effectiveness and Contagiousness of Asymptomatic Individuals: Cross-sectional Survey of Deaf and Hard of Hearing Adults in the United States

Raylene Paludneviciene[1], PhD; Tracy Knight[2], EdD; Gideon Firl[3], BA; Kaela Luttrell[3]; Kota Takayama[4], PhD; Poorna Kushalnagar[1,3], PhD

[1]Department of Psychology, Gallaudet University, Washington, DC, United States

[2]Department of World Languages and Cultures, Sam Houston State University, Huntsville, TX, United States

[3]Center for Deaf Health Equity, Gallaudet University, Washington, DC, United States

[4]Department of Social Work, Gallaudet University, Washington, DC, United States

Corresponding Author:
Poorna Kushalnagar, PhD
Department of Psychology
Gallaudet University
800 Florida Ave NE
Washington, DC, 20002
United States
Phone: 1 585 666 0818
Email: poorna.kushalnagar@gallaudet.edu

Abstract

Background: During the COVID-19 pandemic, there has been a rapid increase in the amount of information about the disease and SARS-CoV-2 on the internet. If the language used in video messages is not clear or understandable to deaf and hard of hearing (DHH) people with a high school degree or less, this can cause confusion and result in information gaps among DHH people during a health emergency.

Objective: The aim of this study is to investigate the relationship between DHH people's perception of the effectiveness of physical distancing and contagiousness of an asymptomatic person.

Methods: This is a cross-sectional survey study on DHH people's perceptions about COVID-19 (N=475). Items pertaining to COVID-19 knowledge were administered to US deaf adults from April 17, 2020, to May 1, 2020, via a bilingual American Sign Language/English online survey platform.

Results: The sample consisted of 475 DHH adults aged 18-88 years old, with 74% (n=352) identifying as White and 54% (n=256) as female. About 88% (n=418) of the sample felt they knew most things or a lot about physical distancing. This figure dropped to 72% (n=342) for the question about the effectiveness of physical distancing in reducing the spread of COVID-19 and 70% (n=333) for the question about the contagiousness of an infected person without symptoms. Education and a knowledge of the effectiveness of physical distancing significantly predicted knowledge about the contagiousness of an asymptomatic individual. Race, gender, and age did not emerge as significant predictors.

Conclusions: This results of this study point to the strong connection between education and coronavirus-related knowledge. Education-related disparities can be remedied by making information fully accessible and easily understood during emergencies and pandemics.

(J Med Internet Res 2021;23(2):e21103) doi:10.2196/21103

KEYWORDS

COVID-19; coronavirus; physical distancing; asymptomatic individual; social media; deaf; hard of hearing; sign language; perception; misinformation

Introduction

Health disparities have been experienced by deaf and hard of hearing (DHH) people in general, as well as during the COVID-19 pandemic. DHH people often do not have equitable access to health information, especially during emergencies [1-3], which further contributes to low perceived quality-of-life outcomes and associated health disparities [4]. For DHH people who speak American Sign Language (ASL), equitable access to health information requires an interpreter to translate between spoken English and ASL, as well as captioning, particularly for information given during live broadcasts about an emergency. To date, not all COVID-19–related video media have included sign language interpreters. However, most DHH ASL speakers, particularly DHH people with a high school degree or less, depend on information presented in ASL in videos as their go-to source for health information, with Facebook being the most often used source [5]. Previous studies have shown that DHH people with a high school degree or less have lower human papillomavirus knowledge, genetic testing awareness, and HIV screening uptake, all of which can be remedied with greater accessibility of pertinent information in ASL [6-8].

While the American public tends to rely on social media to keep up with news [9], social media can also be used as an informal source of information about others' experiences with the health care system or as a means to gain knowledge about specific health information shared by individuals. However, when health information is reproduced and shared informally without oversight by experts, misinformation can occur, including sharing of misinformation related to the ineffectiveness of certain social behaviors in reducing the spread of infections (eg, presenting inaccurate evidence against the use of masks). People who demonstrate low electronic health (eHealth) literacy often fall victim to misinformation as they are not able to verify the accuracy of health information. They may also dismiss information about the severity of certain diseases that was shared by reliable sources on the internet [10-13]. Although the prevalence of misinformation about health on social media is harmful, it can be difficult to correct this misinformation [14]. Consequently, DHH people who have low eHealth literacy and rely on social media to learn information about COVID-19 may be at risk for believing misinformation regarding health and health risks.

During the COVID-19 pandemic, there has been a rapid increase in the amount of information about COVID-19 on the internet, resulting in an "infodemic." This infodemic makes the identification of accurate COVID-19 information from reliable sources difficult [15-17]. DHH people who do not have the requisite background and training in emergency preparedness may select certain information to translate into ASL and share these ASL videos on social media for other DHH people to watch. If the content relayed in these videos is inaccurate or the language used is not clear or understandable to DHH people with a high school degree or less, it can cause confusion and result in information gaps among DHH people during a health emergency. These information gaps can cause harm in medically underserved DHH groups (eg, low education, low income, and low literacy) that are already at risk for disparate outcomes [18]. Accurate critical health information may also not be adequately disseminated to DHH people whose primary language is ASL. These concerns suggest a need to gather a baseline of DHH people's awareness of and risk perceptions related to the current pandemic. Specific research aims for this study include investigating the following: (1) the relationships between DHH people's awareness of physical distancing and perception of the effectiveness of physical distancing in reducing the spread of COVID-19, and (2) the relationship between DHH people's perceived effectiveness of physical distancing and perceived contagiousness of an asymptomatic person.

Methods

A COVID-19 knowledge and risk perception survey in ASL was created by a team of experts in DHH health, translation, and survey development [19,20]. The web link to participate in this survey was disseminated via social media as well as through direct email invitations to those who previously took surveys through the Center for Deaf Health Equity at Gallaudet University.

Data Collection Procedure

Following Institutional Review Board approval by Gallaudet University (RB-FY20-80), research staff recruited DHH individuals throughout the United States, both on social media and through email invitations to past survey participants. Participation in this study required access to a computer and the internet to complete an online bilingual ASL/English survey about knowledge of physical distancing and asymptomatic individuals' contagiousness. As a result, the findings are limited to DHH people with access to computer technology and internet/Wi-Fi connectivity. Only those who self-reported using ASL as their primary language were included because this group was identified as a medically underserved group [21-23]; exclusion criteria included those aged <18 years and those who lost their hearing ability due to aging. After the participant viewed the study information in ASL and English, they were directed to a page where they could choose to provide consent or decline to participate. Following consent, the online survey presented in both ASL and English took approximately 10 minutes to complete. No names or identifying information were included in the survey and a unique identifier was used to avoid storing personal information in the survey data set.

Participant Sources

Participants for this study were drawn from two sources: anonymous participation through a survey link shared via Facebook and a recruitment database pool that was sent an invitation email to take the online survey. The COVID-19 knowledge and risk perception survey included questions about physical distancing awareness, knowledge of physical distancing effectiveness, and knowledge of the contagiousness of an infected person without symptoms. The survey was administered to US Deaf adults from April 17, 2020, to May 1, 2020. Prior to survey administration, all items were translated into ASL with coaching from a Deaf researcher with expertise in translating survey items; the final translations were captured on film.

Survey Items

For the purposes of this study, awareness of physical distancing and perceived effectiveness of physical distancing were assessed with the following questions: "How much do you know about physical distancing?" and "How effective do you think physical distancing is for the prevention of the spread of the coronavirus (in other words: to what extent does your physical distancing behavior contribute to other people not getting sick)?" The response options for the former question were "never knew," "little," "some," "most," and "know a lot," while the response options for the latter question were "not at all effective," "little effective," "somewhat effective," "mostly effective," and "very effective." Participants' perception of the ability of an infected person without symptoms to spread COVID-19 was assessed with the following question: "Based on what you know, how contagious is someone who has been infected with the coronavirus but who shows no symptoms (no coughing; no fever)?" For this question, the response options were "not at all contagious," "little contagious," "somewhat contagious," "contagious," and "very contagious."

Statistical Analyses

Weighted descriptive statistics, such as cross-tabulation and percentage procedures, were used to summarize the sample. As a logistic regression model required binary response variables, responses to the perceived effectiveness of physical distancing and perceived contagiousness of an infected person without symptoms questions were recoded into binary low and high groups. Response options of "not at all effective," "little effective," and "somewhat effective" were recoded into the low group, while "mostly effective" and "very effective" were recoded into the high group. Response options of "not at all contagious," "little contagious," and "somewhat contagious" were recoded into the low group, while "contagious" and "very contagious" were recoded into the high group. A backward elimination variable selection procedure was performed on a logistic regression model, using age, gender, race, education, and baseline physical distancing awareness to predict the odds of having an accurate perception of the effectiveness of physical distancing in reducing the spread of COVID-19 (reference group). The criteria for retaining predictors was a P value <.05. Finally, all variables including physical distancing awareness and knowledge of physical distancing effectiveness were entered as explanatory sets to assess their relationships with knowledge of the contagiousness of an infected person without symptoms (high knowledge of the level of contagiousness was the reference). For observational, cross-sectional studies, a minimum sample size of 400 is recommended for analyses with six predictor variables in the final model [24]. SPSS (Version 25.0; IBM Corp) was used for all analyses.

Results

The weighted sample of those that answered all questions for this study consisted of 475 adults aged 18-88 years old, with 74% (n=352) identifying as White and 54% (n=256) as female (Table 1). When asked about their preferred language on a daily basis (response options: ASL, English, both ASL and English), half of the sample (50%) preferred ASL only. About 88% (n=418) of the sample felt they knew most things about or a lot about physical distancing. This figure dropped to 72% (n=342) for the question about the effectiveness of physical distancing in reducing the spread of COVID-19 and 70% (n=333) for the question about the contagiousness of an infected person without symptoms. Bivariate correlation analysis indicated positive associations among physical distancing awareness, knowledge of physical distancing effectiveness, and knowledge of contagiousness variables (r=0.24-0.38; P<.001).

Table 1. Weighted sample characteristics and logistic regression of perceived contagiousness in asymptomatic individuals.

Sample characteristics	Participants, n (%)	Adjusted odds ratio (95% CI)	*P* value
Age group (years)			.31
18-34	113 (23.8)	1.00	N/A[a]
35-49	154 (32.4)	0.65 (0.29-1.43)	N/A
50-64	126 (26.5)	0.45 (0.20-1.01)	N/A
65-74	58 (12.3)	0.78 (0.29-2.12)	N/A
≥75	24 (5.0)	1.17 (0.23-5.97)	N/A
Gender identity			.22
Male	203 (42.8)	1.00	N/A
Female	256 (53.8)	1.63 (0.92-2.87)	N/A
Nonbinary	16 (3.4)	2.06 (0.24-17.94)	N/A
Race/ethnicity			.74
White	352 (74.2)	1.00	N/A
Black	21 (4.4)	0.61 (0.21-1.79)	N/A
Latinx	67 (14.2)	0.92 (0.34-2.49)	N/A
Asian/Other	35 (7.3)	1.26 (0.53-3.02)	N/A
Education			.001
High school degree	176 (37.0)	1.00	N/A
Some college	175 (36.8)	1.85 (0.80-4.30)	N/A
College degree	124 (26.2)	4.67 (2.22-9.80)	N/A
Physical distancing awareness	N/A	1.66 (0.64-4.32)	.30
Physical distancing effectiveness	N/A	3.13 (1.78-5.50)	.001

[a]N/A: not applicable.

In the first model, a backward elimination variable selection procedure was used to identify the significant predictors for perceiving physical distancing as effective in reducing the spread of COVID-19. In this model, physical distancing awareness and education emerged as significant predictors of perceiving physical distancing as effective (adjusted odds ratio for physical distancing awareness=5.00, 95% CI 2.09-11.95; adjusted odds ratio for education=1.89, 95% CI 1.13-3.16). Race, gender, and age did not predict perceptions of the effectiveness of physical distancing. In the second model, both physical distancing awareness and perception of the effectiveness of physical distancing were entered as a set in the first block. All demographics were entered as a set in the second block. As shown in Table 1, with adjustments for correlates, the results showed that only perception of the effectiveness of physical distancing and education were significant in their relationships with perceived contagiousness of asymptomatic individuals. Compared to DHH respondents who had 12 years of education or lower, DHH respondents with a college degree were nearly five times more likely to report that asymptomatic individuals are contagious (adjusted odds ratio=4.67, 95% CI 2.22-9.80). Again, age, gender, and race did not predict knowledge of contagiousness of an infected person without symptoms.

Discussion

Principal Findings

A majority of this US DHH study sample had heard of physical distancing, but not all believed that physical distancing was effective in reducing the spread of COVID-19 or that an asymptomatic individual was contagious. Above and beyond sociodemographic variables and physical distancing awareness, DHH people who had a college degree were almost twice as likely to believe that physical distancing was effective at reducing the spread of COVID-19 compared to DHH people who did not have a college degree. Although both education and physical distancing awareness strongly predicted perceptions of the effectiveness of physical distancing, this study did not find evidence for the contribution of age, gender, and race to the development of appropriate perception of physical distancing effectiveness. This finding is consistent with research studies that reported a connection between more years of education and a greater level of health-related awareness or screening uptake [6-8]. Our study also showed that DHH people who have a college degree, as well as knowledge of the effectiveness of physical distancing, were much more likely to believe that an infected person without symptoms is contagious compared to DHH people who did not have a college degree and

demonstrated low knowledge of the effectiveness of physical distancing.

The lower knowledge of physical distancing and COVID-19 among DHH people who do not have a college degree may be in part explained by inadequate, inaccessible, or misleading information presented in ASL on social media. According to the 2020 World Health Organization situation report on the novel coronavirus, the COVID-19 outbreak has been accompanied by a global infodemic on social media. As a result, some social media sources may be of low quality, which can potentially lead to the sharing of unreliable information by people in the community. At the same time, social media has been cited as an essential tool for health information seeking among DHH people who do not have a college degree [5] and it can reliably be used to promote rational social health behavior during emergencies and pandemics. Thus, for all social media postings, ASL videos about health during a pandemic must actively cite and include trusted sources for further information. Given the strong connection between education and eHealth literacy [25], as well as the importance of the readability of online health content to accommodate people with a low level of education [26], our community has a responsibility to ensure that social media health videos are fully accessible in ASL and understood by all, including those who have only a high school degree. Clinicians have a responsibility to incorporate patient-centered care and communication in their practice to ensure that patients who are DHH understand what they need to do to take care of their health. Next, we provide recommendations on the development of videos to help increase ASL comprehension for DHH people with a high school education or less.

DHH people with limited access to written English due to language delay or deprivation and DHH immigrants whose first language is not English often have limited ASL fluency [27]. To best accommodate the language needs of these underserved subgroups, ASL videos should be screened for readability for DHH with a high school education or less. This may include avoiding relying on fingerspelling to describe concepts, and instead using visual descriptions and illustrations to make the information easier to understand. Specific concepts might need more elaboration, such as how germs are spread and how masks help prevent the spread of germs. Increasing the use of visual descriptions and elaborating on complex concepts would be key components in providing clear and well-conceived emergency messages. Videos in ASL that include visual descriptions and elaboration of concepts are likely to support comprehension, be effectively transmitted, and ultimately help ensure public health safety by reducing the spread of COVID-19.

According to the National Association of the Deaf Position Statement on Accessible Emergency Management for Deaf and Hard of Hearing People [28], the best practices to disseminate awareness training to DHH citizens during the pandemic are the following: (1) ensuring that the materials are accessible in a language that they understand, (2) including medical and public health experts who use ASL in interactive panels, (3) providing training to qualified sign language interpreters, and (4) ensuring that all news and notification systems are accessible by any form of telecommunication technology.

Conclusion

The results of this study point to the strong connection between education and COVID-19–related knowledge. Therefore, the information that DHH organizations and public health agencies disseminate during emergencies and pandemics must be clear, contain adequate and reliable information, and be timely in concordance with other information being disseminated. For online streaming of time-sensitive news, such as state and local government updates, live captioning in addition to sign language interpreting would provide DHH people with full access to information. ASL videos that are created for social media or the internet must include features that support ease of understanding, which fosters improved knowledge and more accurate perceptions. Information in ASL disseminated through other sources (such as television and the news) must also be clear and accessible to the whole DHH community.

Acknowledgments

This work was supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) of the National Institutes of Health (R01DC014463-01A1 and R01DC014463-05S4 to Poorna Kushalnagar, PhD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Takayama K. Disaster Relief and Crisis Intervention with Deaf Communities: Lessons Learned from the Japanese Deaf Community. J Soc Work Disabil Rehabil 2017;16(3-4):247-260 [FREE Full text] [doi: 10.1080/1536710X.2017.1372241] [Medline: 28976290]
2. Neuhauser L, Ivey SL, Huang D, Engelman A, Tseng W, Dahrouge D, et al. Availability and readability of emergency preparedness materials for deaf and hard-of-hearing and older adult populations: issues and assessments. PLoS One 2013;8(2):e55614 [FREE Full text] [doi: 10.1371/journal.pone.0055614] [Medline: 23451029]

3. Ivey SL, Tseng W, Dahrouge D, Engelman A, Neuhauser L, Huang D, et al. Assessment of state- and territorial-level preparedness capacity for serving deaf and hard-of-hearing populations in disasters. Public Health Rep 2014 Mar;129(2):148-155 [FREE Full text] [doi: 10.1177/003335491412900208] [Medline: 24587549]
4. Kushalnagar P, McKee M, Smith SR, Hopper M, Kavin D, Atcherson SR. Conceptual model for quality of life among adults with congenital or early deafness. Disabil Health J 2014 Jul;7(3):350-355 [FREE Full text] [doi: 10.1016/j.dhjo.2014.04.001] [Medline: 24947577]
5. Kushalnagar P, Kushalnagar R. Health-related Information Seeking among Deaf Adults: Findings from the 2017 Health Information National Trends Survey in American Sign Language (HINTS-ASL). In: Hale T, Wen-Ying SC, Cotten S, Khilnani A, editors. eHealth: Current Evidence, Promises, Perils and Future Directions. United Kingdom: Emerald Insights; Aug 06, 2018.
6. Spellun AH, Moreland CJ, Kushalnagar P. Young Deaf Adults' Knowledge of Human Papillomavirus and Human Papillomavirus Vaccine's Effectiveness in Preventing Cervical, Anal, Penile, and Oral Cancer. J Pediatr Adolesc Gynecol 2019 Jun;32(3):293-299 [FREE Full text] [doi: 10.1016/j.jpag.2018.11.013] [Medline: 30529699]
7. Kushalnagar P, Holcomb J, Sadler GR. Genetic testing and eHealth usage among Deaf women. J Genet Couns 2019 Oct;28(5):933-939 [FREE Full text] [doi: 10.1002/jgc4.1134] [Medline: 31180177]
8. Argenyi M, Kushalnagar P. Social Media Use and HIV Screening Uptake Among Deaf Adults in the United States: Cross-Sectional Survey Study. JMIR Public Health Surveill 2019 Oct 02;5(4):e13658 [FREE Full text] [doi: 10.2196/13658] [Medline: 31579021]
9. Mitchell A, Gottfried J, Kiley J, Matsa KE. Political Polarization and Media Habits. Pew Research Center Journalism & Media. 2014 Oct 21. URL: https://www.journalism.org/2014/10/21/political-polarization-media-habits/ [accessed 2020-05-30]
10. Bode L, Vraga EK. See Something, Say Something: Correction of Global Health Misinformation on Social Media. Health Commun 2018 Sep;33(9):1131-1140. [doi: 10.1080/10410236.2017.1331312] [Medline: 28622038]
11. Bode L, Vraga EK. In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media. J Commun 2015 Jun 23;65(4):619-638. [doi: 10.1111/jcom.12166]
12. Tan ASL, Lee C, Chae J. Exposure to Health (Mis)Information: Lagged Effects on Young Adults' Health Behaviors and Potential Pathways. J Commun 2015 Jul 06;65(4):674-698. [doi: 10.1111/jcom.12163]
13. Jolley D, Douglas KM. The effects of anti-vaccine conspiracy theories on vaccination intentions. PLoS One 2014 Feb 20;9(2):e89177 [FREE Full text] [doi: 10.1371/journal.pone.0089177] [Medline: 24586574]
14. Nyhan B, Reifler J, Richey S, Freed GL. Effective messages in vaccine promotion: a randomized trial. Pediatrics 2014 Apr;133(4):e835-e842. [doi: 10.1542/peds.2013-2365] [Medline: 24590751]
15. Hao K, Basu T. The coronavirus is the first true social-media 'infodemic'. MIT Technol Rev 2020. 2020 Feb 12. URL: https://www.technologyreview.com/2020/02/12/844851/the-coronavirus-is-the-first-true-social-media-infodemic/ [accessed 2020-05-30]
16. Lai I, Gupta S. Non-traditional outbreak surveillance - the 15 year experience. International Journal of Infectious Diseases 2019 Feb;79:114. [doi: 10.1016/j.ijid.2018.11.282]
17. Anzar W, Baig QA, Afaq A, Taheer TB, Amar S. Impact of Infodemics on Generalized Anxiety Disorder, Sleep Quality and Depressive Symptoms among Pakistani Social Media Users during Epidemics of COVID-19. Merit Res J Med Med Sci 2020 Mar 25:1-5. [doi: 10.5281/zenodo.3727246]
18. Schild S, Dalenberg CJ. Trauma exposure and traumatic symptoms in deaf adults. Psychological Trauma: Theory, Research, Practice, and Policy 2012;4(1):117-127. [doi: 10.1037/a0021578]
19. Kushalnagar P, Paludneviciene R, Kallen M, Atcherson S, Cella D. PROMIS-deaf profile measure: cultural adaptation and psychometric validation in American sign language. J Patient Rep Outcomes 2020 Jun 09;4(1):44 [FREE Full text] [doi: 10.1186/s41687-020-00208-7] [Medline: 32519000]
20. Kushalnagar P, Harris R, Paludneviciene R, Hoglind T. Health Information National Trends Survey in American Sign Language (HINTS-ASL): Protocol for the Cultural Adaptation and Linguistic Validation of a National Survey. JMIR Res Protoc 2017 Sep 13;6(9):e172 [FREE Full text] [doi: 10.2196/resprot.8067] [Medline: 28903891]
21. Smith S, Chin N. Social Determinants of Health in Deaf Communities. In: Public Health - Social and Behavioral Health. London, UK: InTechOpen Limited; 2012:449-460.
22. Barnett S, McKee M, Smith S, Pearson T. Deaf sign language users, health inequities, and public health: opportunity for social justice. Prev Chronic Dis 2011 Mar;8(2):A45 [FREE Full text] [Medline: 21324259]
23. McKee MM, Paasche-Orlow MK, Winters PC, Fiscella K, Zazove P, Sen A, et al. Assessing Health Literacy in Deaf American Sign Language Users. J Health Commun 2015;20 Suppl 2:92-100 [FREE Full text] [doi: 10.1080/10810730.2015.1066468] [Medline: 26513036]
24. Bujang MA, Sa'at N, Sidik TMITAB, Joo LC. Sample Size Guidelines for Logistic Regression from Observational Studies with Large Population: Emphasis on the Accuracy Between Statistics and Parameters Based on Real Life Clinical Data. Malays J Med Sci 2018 Jul 30;25(4):122-130 [FREE Full text] [doi: 10.21315/mjms2018.25.4.12] [Medline: 30914854]
25. Knapp C, Madden V, Wang H, Sloyer P, Shenkman E. Internet use and eHealth literacy of low-income parents whose children have special health care needs. J Med Internet Res 2011 Sep 29;13(3):e75 [FREE Full text] [doi: 10.2196/jmir.1697] [Medline: 21960017]

26. Kim H, Xie B. Health literacy in the eHealth era: A systematic review of the literature. Patient Educ Couns 2017 Jun;100(6):1073-1082. [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]
27. Vernon M, Raifman LJ, Greenberg SF, Monteiro B. Forensic pretrial police interviews of deaf suspects avoiding legal pitfalls. International Journal of Law and Psychiatry 2001 Jan;24(1):43-59. [doi: 10.1016/s0160-2527(99)00031-x]
28. Position Statement on Accessible Emergency Management for Deaf and Hard of Hearing People. National Association of the Deaf. 2020. URL: https://www.nad.org/about-us/position-statements/position-statement-on-accessible-emergency-management-for-deaf-and-hard-of-hearing-people/ [accessed 2020-09-28]

Abbreviations

ASL: American Sign Language
DHH: deaf and hard of hearing
eHealth: electronic health

Edited by C Basch; submitted 05.06.20; peer-reviewed by M Argenyi, A Jaiswal; comments to author 21.08.20; revised version received 30.09.20; accepted 07.02.21; published 25.02.21.

Please cite as:
Paludneviciene R, Knight T, Firl G, Luttrell K, Takayama K, Kushalnagar P
Perception of COVID-19 Physical Distancing Effectiveness and Contagiousness of Asymptomatic Individuals: Cross-sectional Survey of Deaf and Hard of Hearing Adults in the United States
J Med Internet Res 2021;23(2):e21103
URL: https://www.jmir.org/2021/2/e21103
doi:10.2196/21103
PMID:33560996

Original Paper

Impact of Remote and Virtual Care Models on the Sustainability of Small Health Care Businesses: Perceptual Analysis of Small Clinics, Physician Offices, and Pharmacies in Colorado

Madhavan Parthasarathy[1], MBA, PhD; Jiban Khuntia[1], PhD; Rulon Stacey[1], MHA, PhD

University of Colorado-Denver, Denver, CO, United States

Corresponding Author:
Madhavan Parthasarathy, MBA, PhD
University of Colorado-Denver
1475 Lawrence Street, Denver, CO 80202
Denver, CO
United States
Phone: 1 303 315 8445
Email: madhavan.parthasarathy@ucdenver.edu

Abstract

Background: Lockdowns and shelter-in-place orders during COVID-19 have accelerated the adoption of remote and virtual care (RVC) models, potentially including telehealth, telemedicine, and internet-based electronic physician visits (e-visits) for remote consultation, diagnosis, and care, deterring small health care businesses including clinics, physician offices, and pharmacies from aligning resources and operations to new RVC realities. Current perceptions of small health care businesses toward remote care, particularly perceptions of whether RVC adoption will synergistically improve business sustainability, would highlight the pros and cons of rapidly adopting RVC technology among policy makers.

Objective: This study aimed to assess the perceptions of small health care businesses regarding the impact of RVC on their business sustainability during COVID-19, gauge their perceptions of their current levels of adoption of and satisfaction with RVC models and analyze how well that aligns with their perceptions of the current business scenario (SCBS), and determine whether these perceptions influence their view of their midterm sustainability (SUST).

Methods: We randomly sampled small clinics, physician offices, and pharmacies across Colorado and sought assistance from a consulting firm to collect survey data in July 2020. Focal estimated study effects were compared across the three groups of small businesses to draw several insights.

Results: In total, 270 respondents, including 82 clinics, 99 small physician offices, and 89 pharmacies, across Colorado were included. SRVC and SCBS had direct, significant, and positive effects on SUST. However, we investigated the effect of the interaction between SRVC and SCBS to determine whether RVC adoption aligns with their perceptions of the current business scenario and whether this interaction impacts their perception of business sustainability. Effects differed among the three groups. The interaction term SRVC×SCBS was significant and positive for clinics (P=.02), significant and negative for physician offices (P=.05), and not significant for pharmacies (P=.76). These variations indicate that while clinics positively perceived RVC alignment with the current business scenario, the opposite held true for small physician offices.

Conclusions: As COVID-19 continues to spread worldwide and RVC adoption progresses rapidly, it is critical to understand the impact of RVC on small health care businesses and their perceptions of long-term survival. Small physician practices cannot harness RVC developments and, in contrast with clinics, consider it incompatible with business survival during and after COVID-19. If small health care firms cannot compete with RVC (or synergistically integrate RVC platforms into their current business practices) and eventually become nonoperational, the resulting damage to traditional health care services may be severe, particularly for critical care delivery and other important services that RVC cannot effectively replace. Our results have implications for public policy decisions such as incentive-aligned models, policy-initiated incentives, and payer-based strategies for improved alignment between RVC and existing models.

(J Med Internet Res 2021;23(2):e23658) doi:10.2196/23658

KEYWORDS

clinics; Colorado; COVID-19; electronic physician visits; entrepreneur; pharmacies; physician offices; small business; telehealth; telemedicine

Introduction

This study assessed the perceptions of small health care business owners toward remote and virtual care (RVC) services, which include telemedicine, telehealth, electronic physician visits, and internet-mediated communications and have become increasingly prevalent during COVID-19 with lockdowns, shelter-in-place orders, and other social distancing measures [1-4], effectively restricting patient access to in-person care. Consequently, in-person care–based small businesses have been severely impacted.

Hospitals and health systems are currently attempting to modify their care delivery methods in response to COVID-19. They have used various measures, such as discontinuing all elective procedures, converting ambulatory surgery suites to intensive care units, and moving a large portion of their supply offline to generate capacity in the event that the volume of patients with COVID-19 exceeds the projected volume. However, while hospitals and health systems have received financial incentives from the Federal Emergency Management Agency during COVID-19, no such incentives have been provided to the more than half a million small health care businesses, clinics, physician offices, blood banks, laboratories, and pharmacies, which form the traditional core of the health care ecosystem in the United States [5,6]. Consequently, many of these traditional small businesses have been experiencing considerable financial distress.

Over the past two decades, the health care industry had experienced a significant influx of entrepreneurs intending to capitalize on the country's expanding health care expenditures, which amount for almost 20% of the gross domestic product, with an average of 5.5% growth projected through 2026 [7]. These entrepreneurs have often found themselves in a position of "doing good" as they provided much-needed services for the aging, ill, and disabled population at convenient locations. With the recent increase in the need for drug and substance abuse services and recent advancements in medical and technological treatments, they have entered a valuable niche in the health care market. Considering this trend in the health care industry, a plethora of businesses have been operating in the United States under the umbrella term of "clinics" to cater to the differing needs of care of the population. These clinics have been providing affordable care in areas such as physical and occupational therapy, diabetes care, kidney care, drug treatment, and rehabilitation centers, alternate medicine, hearing aid clinics, imaging centers, and nutrition or dietician services [8,9].

Similarly, small physician offices (ie, those with <5 physicians) constituted a significant component of the US health care system, garnering 54.5% of the total of 883.7 million visits to physician offices in 2016 [10]. Press reports reveal that small physician offices are under distress during COVID-19, and that "that won't pay the light bill or the rent" [11], even though they note that these businesses are ready to respond to the challenge and provide care during the pandemic if they can receive incentives to stay operational.

Independent pharmacies in the United States have also been recently struggling, with their market share declining by approximately 10% in the last decade, although they still account for 35% of all retail pharmacies and a US $77.6 billion market share in 2017 [12]. Recent efforts to enhance price transparency have alleviated the "gag clauses," which have served as a primary contributor to high out-of-pocket expenditures for drugs for patients [13,14] and have also resulted in a commensurate reduction in profit margins for some of those pharmacies. Stricter regulatory control has decreased the availability of over-the-counter medication now rather than a decade ago. Furthermore, physicians affiliated with health systems often electronically prescribe medication to be obtained from large pharmacy chains affiliated with the health care system, which have proliferated wider than before and are reportedly diverting large chunks of revenue from small, independently owned pharmacies [15].

Considering the lockdowns and shelter-in-place directives during the pandemic, consumers not only have greater access to web-based physician consultations but also have patronized online pharmacies [16]. Consequently, traditional pharmacies are losing their market share, leading to severe financial strain on them [17]. Indeed, more than half of the independent pharmacy owners are experiencing a negative revenue influx [18], with fewer individuals visiting stores [19]. Unlike large pharmaceutical companies, small pharmacies have also struggled to embrace the emerging RVC models and be sustainable, and this remains a concern [20].

This leads to the question of why these small businesses do not adopt RVC. We offer several conjectures to answer this question. First, RVC might be perceived as a temporary phenomenon, which could decline post COVID-19 because an online consultation cannot effectively replace an in-person consultation in many circumstances. Second, there may be a genuine fear that resource limitations may inhibit the ability of these small businesses to adopt RVC, and this may deter their sustained survival. Costs associated with digital infrastructure investment, platforms, connections, training, test runs, and scheduling services may be perceived as prohibitive. Third, whether organizations and insurance companies can reimburse RVC models remains a concern. Finally, assessment of the costs and benefits of adopting an RVC model can be difficult. The potentially expensive and arduous process of adopting an RVC model remains a major task, and there is no clarity on whether all patients and organizations would eventually transition to these models. Maintaining a brick-and-mortar store along with the RVC may be costly and overwhelming to these small business entities.

Small businesses can certainly benefit from integrating RVC processes into their existing business operations. Consumers, both on the demand and supply sides of care, are familiar with

the use of cell phones, video chat platforms, text messaging, and cloud-based communication; hence, they can adopt a digital care platform without adequate training. Further, RVC models help overcome physical barriers, thereby increasing the reach of small businesses (and competition) to a much larger geographic area. Data collected by RVC systems may help identify how small firms may gain more business, connect patients with specialists in just-in-time mode, and improve branding.

This study aims to address these issues through a perceptual survey of small physician-owned businesses, clinics, and pharmacies. We sought to determine whether the owners of these businesses view RVC as a threat, whether they consider themselves capable of integrating the many opportunities accompanying RVC with their existing business model, and whether they fear midterm survival in the wake of RVC.

To answer these questions, we propose three conjectures based on our expectation that clinics, as opposed to small physician-owned businesses and pharmacies, will be better able to realize and leverage the value proposition of RVC because many clinics are affiliated with larger organizations that already have established digital platforms. Some of these may be retail or franchise establishments, perhaps connected to a larger health care system or hospital, which renders them more likely to be able to leverage the organization's existing RVC or other digital platforms to their benefit to a larger extent than independent physician-owned businesses or pharmacies. Our conjectures are the following:

1. Small clinic owners will perceive value in aligning opportunities accompanying RVC models with their current business practices.
2. Small physician offices will not perceive value in aligning opportunities accompanying RVC models with their current business practices.
3. Small and independent pharmacy owners will not perceive value in aligning opportunities accompanying RVC models with their current business practices.

Methods

Recruitment

We conducted a focus group study with 8 health care experts via Zoom; they comprised 4 entrepreneurs familiar with the health care environment in Colorado, 2 physicians, and 2 executive health faculty members. The discussion centered around the impact of RVC on the sustainability of small physician offices, clinics, and pharmacies during COVID-19. The experts generally supported the notion that RVC adoption is posing a challenge to small health care businesses. As conjectured, the experts believed that clinics, by their scale and scope, would likely not only benefit from but also be able to rapidly integrate RVC into their business practices, more so than physicians and pharmacies. Furthermore, the experts believed that physician offices are most likely to directly compete with, and hence be threatened by, RVC. Clinics often provide in-person consultations; therefore, the focus group speculated that RVC was not a major direct threat to their sustainability as it was for physician offices. Furthermore, the experts speculated that small local pharmacies would be negatively impacted during COVID-19 and that consumers would increasingly tend to visit online pharmacies, thereby avoiding waiting in line at physical locations that may increase the risk of infection. To investigate whether these opinions were reflective of health care businesses in Colorado, our study was aimed at empirically validating these presumptions, using a scientifically collected sample of small health care businesses in Colorado.

Data Collection

Considering the time sensitivity of this study, we solicited a professional consulting firm for data collection. The consulting firm sampled 445 small health care businesses in Colorado (135 clinics, 141 physician offices, and 169 pharmacies) through an opt-in approach during June-July 2020. Of these, 282 firms responded to the survey. In total, 12 observations were excluded owing to missing responses, resulting in a final sample size 270 businesses consisting of 82 clinics, 99 physician offices, and 89 pharmacies. The identities of the respondents were not disclosed.

The survey instrument used to collect the data is presented in Table 1. Questions regarding the current sustainability of the business in the subsequent 1-3 years were asked to the survey participants. Furthermore, satisfaction with the current state of the business and current remote care practices, if any, adopted by the business were measured. The survey was brief and succinct, taking only 5 minutes to complete, addressing only critical research questions and eliciting realistic responses.

The survey instrument, consisting of only 11 questions, was piloted with a sample of 22 respondents, leading to minor adjustments to a few items. Responses were coded, validated, and analyzed using STATA (version 14.2, StataCorp).

Table 1. Survey questionnaire and coding scheme.

Variable	Description	Questionnaire items	References
Dependent variables			
Sustainability of the physician office, clinic, or pharmacy (SUST)	Business owner's perception of practice or business survival in the near future.	At an overall level, to what extent do you feel your clinic/practice/pharmacy is sustainable, considering the current COVID-19 scenario, in a span of the next 1-3 years?	Scale: 1 (least confident) to 5 (very confident)
Independent variables			
Satisfaction with the current business scenario (SBCS)	Business owner's satisfaction with the current practice and/or business scenario during COVID-19.	Questions: to what extent are you currently [during COVID-19 situation] satisfied with: (1) the overall clinic/practice/pharmacy relevant business scenario, situated across your city/town/county, (2) the current method of practice or business operations, and (3) the quality of day-to-day business transactions?	Scale: 1 (very dissatisfied) to 5 (very satisfied); Cronbach α=.79
Satisfaction with remote and virtual care models (including telemedicine, telehealth, and e-visits) (SRVC)	Business owner's satisfaction with the current (during COVID-19) remote care (telemedicine, telehealth, remote consulting, and e-visit provisions) provided by your clinic/practice/pharmacy.	Questions: to what extent are you currently (during COVID-19 situation) satisfied with the provisions and operations in your organization relevant to: (1) remote care, (2) telehealth and telemedicine, and (3) remote consulting and e-visits?	Scale: 1 (very dissatisfied) to 5 (very satisfied); Cronbach α=.82
Control variables			
Age (AGE)	Age of the owner	My age is_____.	N/A[a]
Gender (GEN)	Gender of the owner	I am: (a) male and (b) female	N/A
Income (INC)	Net income of the owner from the business	My annual income from the clinic/practice/pharmacy relevant business: (1) less than US $50,000, (2) US $50,000-US $100,000, (3) US $100,000-US $200,000, (4) US $200,000-US $300,000, (5) US $300,000-US $500,000, (6) US $500,000-US $750,000, (7) US $750,000-US $1,000,000, and (8) higher than US $1,000,000.	N/A
Education (EDU)	Education of the owner	My education is: 1=illiterate, 2=up to middle school, 3=up to high school, 5=some college, 6=undergraduate level, and 7=graduate and above.	N/A

[a]N/A: not applicable.

Sample Descriptors and Demographics

Table 2 shows the descriptive statistics and pairwise correlations among the key variables. Most of the owners (mean 3.09, SD 0.6) believe that their business is on the verge of being sustainable, notwithstanding certain concerns. Owner satisfaction with the current business scenario was moderate to high (mean 3.16, SD 0.67). In contrast, owner satisfaction with remote care was low (mean 2.87, SD 0.76), probably owing to the unpreparedness and uncertainties associated with the delivery and rendering of remote care channels and options.

Of the 270 respondent business owners and chief executives, 229 (85%) identified as male. Respondent ages varied from 25-60 years, with 122 (45%) participants aged 25-35 years, 31 (12%) participants aged 35-45 years, 115 (43%) participants aged 45-55 years, and only 2 (0.01%) participants aged 55-60 years.

Further, 26 (10%) participants reported that they earned less than US $50,000 annually, 34 (13%) earned between US $50,000-US $100,000, 90 (33%) earned between US $100,000-US $200,000, 16 (6%) earned between US $300,000-US $500,000, 13 (5%) earned between US $500,000-US $750,000, 10 (4%) earned between US $750,000-US $1,000,000, and 19 (7%) earned more than US $1,000,000.

In addition, 1 (0.4%) participant reported having no education, 1 (0.4%) participant had 4 years of schooling, 4 (1.5%) participants had middle school education, 18 (7%) participants had high school education, 34 (13%) participants had some college education, 112 (42%) participants had an undergraduate degree, and 100 (37%) participants had graduate-level or higher education.

Table 2. Summary statistics and pairwise correlations among variables (N=270).

#	Variable	Mean (SD)	Minimum	Maximum	1	2	3	4	5	6	7
1	SUST[a]	3.09 (0.60)	1	4	1.00						
2	SCBS[b]	3.16 (0.67)	1	4	0.52	1.00					
3	SRVC[c]	2.87 (0.76)	1	4	0.48	0.31	1.00				
4	Age	37.45 (8.37)	25	60	0.02	0.07	0.06	1.00			
5	Gender	0.85 (0.36)	0	1	0.08	0.03	0.05	0.03	1.00		
6	INC[d]	4.67 (1.82)	2	8	–0.13	–0.14	–0.03	0.03	0.00	1.00	
7	EDU[e]	6.03 (1.02)	1	7	–0.14	–0.10	–0.05	–0.14	0.07	0.17	1.00

[a]SUST: sustainability of the physician office, clinic, or pharmacy.

[b]SCBS: satisfaction with the current business scenario.

[c]SRVC: satisfaction with remote and virtual care models (including telemedicine telehealth and e-visits).

[d]INC: income.

[e]EDU: education.

Study Variables

We used a dependent variable reflecting the owner's perception of the business sustainability of physician offices, clinics, or pharmacies in the near future (SUST). This variable was measured using a single-item question. On average, SUST had a mean score of 3.09 (SD 0.6) of 5, indicating that most businesses are somewhat confident in remaining sustainable (Table 2).

Two independent variables were of interest in this study. First, the independent variable of satisfaction with the current business scenario (SCBS) measured the owner's satisfaction with the current practice or the business scenario during COVID-19. This variable was operationalized with three questions on the business scenario across the city/town/county, the current ways of practice or business operations, and the quality of day-to-day business transactions during the pandemic. The internal consistency of the items was high, as reflected by their Cronbach α of .79 and a variable response of 3.16 (SD 0.67) (Table 2).

Second, the satisfaction with RVC (SRVC) models indicated the owner's or chief executive's satisfaction with current remote care (telemedicine, telehealth, remote consulting, and e-visit provisions) provided by their clinics, practice offices, or pharmacies. This variable was operationalized with using three questions on the provisions and aligned operations in the clinic, physician office, or pharmacy relevant to remote care, telehealth, and telemedicine, and remote consulting during COVID-19 (Table 1). The items had high internal consistency with a Cronbach α of .82 and a variable response of 2.87 (SD 0.76) (Table 2).

We were particularly interested in exploring how owners perceived the importance of remote care to remain competitive and mitigate concerns relevant to their current business scenario during COVID-19. To explore this, we established an interaction variable multiplying SCBS and SRVC (SCBS×SRVC), which was used in the models to reflect the net effect of satisfaction with the current business and existing remote care provisions on perceptions of business sustainability.

In addition to these key variables of interest, a limited set of control variables such as age, gender, income, and education were included to account for potential counterfactual explanations relevant to our models.

Econometric Analysis

The empirical model used herein analyzed the impact of owners' perceptions of their current business scenario and their ensuing use of remote care systems on the perceived sustainability of their small businesses. We used an ordinary least square estimation as follows:

$$Y[SUST]_i = \beta_0 + \beta_1\ SCBS_i + \beta_2\ SRVC_i + \beta_3\ SCBS_i \times SRVC_i + Controls_i + \varepsilon_I \quad (1)$$

where *Y* is the dependent variable; β_0, β_1, β_2, and β_3 are the parameter coefficients; and ε is the disturbance associated with each observation. We estimated this model for all businesses in our sample and then separately for the subsamples of clinics, physician offices, and pharmacies.

Results

Table 3 presents the estimation results obtained using equation (1) and summarizes the estimates for the entire sample and individually for clinics, physician offices, and pharmacies.

We found that both SCBS and SRVC variables have significant, positive direct effects on SUST, which indicates that the effort invested in the current situation and in adopting remote care systems influenced the sustainability of the businesses as perceived by the owners. However, this study aimed to investigate whether remote care provision was perceived as adequate in the current business scenarios to influence business sustainability. Accordingly, the coefficients of the interaction variable SCBS×SRVC were further analyzed.

Table 3. Estimation results (N=270).

Variables	Clinics		Physician offices		Pharmacies		Full sample	
	SUST[a]	*P* value	SUST	*P* value	SUST	*P* value	SUST	*P* value
SCBS[b], mean (SE)	0.556 (0.094)	<.001	0.187 (0.097)	.06	0.394 (0.101)	<.001	0.352 (0.056)	<.001
SRVC[c], mean (SE)	0.235 (0.070)	.001	0.159 (0.076)	.04	0.351 (0.070)	<.001	0.277 (0.041)	<.001
SCBS×SRVC, mean (SE)	0.274 (0.110)	.02	-0.182 (0.092)	.05	0.026 (0.083)	.76	–0.011 (0.052)	.83
Age, mean (SE)	0.000 (0.006)	.95	–0.000 (0.007)	.94	–0.001 (0.006)	.81	–0.003 (0.004)	.46
Gender, mean (SE)	–0.028 (0.175)	.87	0.141 (0.137)	.31	0.177 (0.133)	.19	0.091 (0.081)	.26
Income, mean (SE)	–0.027 (0.092)	.77	–0.029 (0.024)	.23	–0.004 (0.026)	.87	–0.017 (0.016)	.31
Education, mean (SE)	–0.059 (0.053)	.27	–0.037 (0.051)	.47	–0.100 (0.050)	.05	–0.048 (0.029)	.10
Constant, mean (SE)	3.721 (0.569)	<.001	3.331 (0.461)	<.001	3.915 (0.366)	<.001	3.626 (0.243)	<.001
Observations, n	82	N/A[d]	99	N/A	89	N/A	270	N/A
R^2	0.477	N/A	0.329	N/A	0.492	N/A	0.393	N/A
Adjusted R^2	0.427	N/A	0.278	N/A	0.448	N/A	0.377	N/A
F test (*df*)	9.638 (7)	N/A	6.383 (7)	N/A	11.200 (7)	N/A	24.280 (7)	N/A
Prob>F	0.000	N/A	0.000	N/A	0.000	N/A	0.000	N/A

[a]SUST: sustainability of the physician office, clinic, or pharmacy.

[b]SCBS: satisfaction with the current business scenario.

[c]SRVC: satisfaction with remote and virtual care models (including telemedicine telehealth and e-visits).

[d]N/A: not applicable.

First, we found that for clinics, the estimation coefficient was positive and significant (β=.274; *P*=.02). This finding indicated a positive interaction between the current business scenario and the adoption of remote care platforms, reflecting higher perceived business sustainability. Second, for physician offices, the opposite trend was observed, in that the estimation coefficient was negative and significant (β=–.182; *P*=.05). This finding indicates a negative interaction between their satisfaction with RVC and their satisfaction with their current business scenario, reflecting lower future sustainability of their businesses. Third, for pharmacies, the estimation coefficient was not significant (β=.026; *P*=.76). This finding indicates that pharmacy owners did not perceive the interaction between their satisfaction of remote care platforms and the current business scenario to be associated with the sustainability of their businesses. Finally, the estimation coefficient of SCBS×SRVC of the complete sample was not significant (*P*=.83), probably owing to an aggregation of positive and negative interactions across the subsamples.

Discussion

Practice and Policy Implications of the Findings

RVC adoption (ie, telemedicine, telehealth, e-visits, and similar internet-enabled treatment options) has grown significantly during COVID-19. RVC fulfills a widespread patient-centric need for access to services of the US health care sector, which is likely to be a high-growth sector in the near future. The growth of RVC has impacted traditional small health care businesses that have traditionally relied on in-person care and serve a local geographic area. Our study was aimed at measuring the perceptions of RVC of these small business owners and determining whether they believe that RVC would threaten their survival or sustainability or, in contrast, if they perceived RVC as a growth opportunity.

Our study assessed the perceptions of three categories of health care providers: physician offices, clinics, and pharmacies. We found that clinic owners were most likely to believe that their adoption of the RVC model interacted positively with their perceptions of the current business scenario, which, in turn, led to a more positive perception of their business sustainability over the subsequent 1-3 years. In other words, their satisfaction with the RVC model was synergistic with their satisfaction with

current business practices, and this was associated with a more optimistic opinion of the firm's midterm business survival. These findings may be certainly moderated by the type of clinic in question. A majority of clinics, such as physical therapy businesses, imaging services, or diabetes care centers that depend on in-person visits, may not perceive RVC as a direct threat. For them, adoption of a digital platform, particularly to enhance efficiency, customer service, and information provision, may complement their services and extend their reach to a younger, more digitally savvy audience.

However, this was not the case with physician offices. The interaction between their satisfaction with the current business scenario and RVC was negative, resulting in a lower perception of their firm's midterm sustainability. In other words, these business owners speculated that RVC, when combined with the existing business scenario, was negatively associated with their likely midterm survival. While the interaction between their perception of the current business scenario and RVC was negative for pharmacies as well, the coefficient was nonsignificant, suggesting that pharmacies did not consider their satisfaction with RVC and the current business scenario to impact their business's midterm sustainability.

Overall, clinics perceived RVC and the current business scenario to be positively associated with business sustainability. In contrast, physician offices speculated that RVC adoption along with the current in-person business scenario was negatively associated with their business's midterm survival.

Many of these small businesses do not qualify for financial assistance from the Federal Emergency Management Agency during COVID-19, and this has significantly, albeit temporarily, decelerated business operations during the pandemic and caused them financial distress. Furthermore, their inability to rapidly adopt a satisfactory and synergistic RVC model to atone for the loss of traditional customers further aggravated this problem. The recent policy-level incentive to render RVC models a permanent provision in the US health care system, without considering the opinions and realities of small health care businesses (which are the backbone of the traditional health care delivery system), seems myopic. RVC cannot replace traditional in-person health care provision in all circumstances, and if physician offices and other small health care businesses cannot survive the competition posed by the RVC model, the results could lead to a marked deficit in critical care. Hence, new technologies should be adopted in a nuanced manner, with policy-level support for these small businesses, thus laying the foundation for gradual and progressive RVC adoption [18].

Certain other issues also need careful policy restructuring: (1) linking insurance payments and payor guarantees to encourage small businesses to adopt the RVC model; (2) developing creative, new consumer-centric models in this context; (3) ensuring affordable RVC platforms for small businesses with reliable technological and infrastructural logistics to ensure that remote care is error-free and of the highest quality; and (4) integrating RVC models with traditional brick-and-mortar health care models that have been established in accordance with the high-touch concept. If remote RVC technology can complement and not compete with in-person treatment models, such a system would have a high combined efficiency. For instance, technologies such as remote patient monitoring and vital reading systems can help improve the impact of traditional systems.

A related implication is to prepare current and next-generation patients and care providers to embrace RVC-related changes that have not been and are not a part of the curriculum in medical or business schools. Medical practitioners may not be good business persons and vice versa. Small physician offices have been particularly faced with this conundrum while facing several regulatory and practice-level changes, which were not a part of their curriculum. Advancements in RVC involve artificial intelligence, machine learning, data science, and above all, a set of customer-oriented, internet-based patient communication and payment or profit models, which need both medical and business skills. The future curriculum should foster the acquisition of business knowledge and skills.

Conclusions

The rapid spread of COVID-19 has yielded unprecedented challenges. Simultaneously, it has helped to rapidly accelerate the adoption of RVC models. RVC serves as a starting point for several innovations that would potentially modify the existing care models and, when used carefully, may lead to a synergistic mixed model of in-person and internet-based care provision. However, our results show that RVC adoption may have progressed too rapidly, with small businesses struggling to integrate RVC technologies with their existing in-person platforms. Physician offices, in particular, are experiencing incompatibility between their current business models and RVC technologies and are concerned about the potential impact of RVC adoption on their midterm survival. If a significant proportion of brick-and-mortar physician offices become nonoperational, the collective impact of such a closure on the US health care system could be catastrophic. Hence, careful consideration of the adoption of RVC models (potentially including business education, government incentives, and support for small businesses) is vital to the sustained adoption and functionality of RVC models. While it is clear that the future of health care depends on these transformative models, the complexity lies in managing these models with appropriate support, training, reimbursement, and incentive models for small businesses, while meeting the expectations and demands of patient-centric health care delivery.

Conflicts of Interest

None declared.

References

1. Fisk M, Livingstone A, Pit SW. Telehealth in the Context of COVID-19: Changing Perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 2020 Jun 09;22(6):e19264 [FREE Full text] [doi: 10.2196/19264] [Medline: 32463377]
2. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine During the COVID-19 Pandemic: Experiences From Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
3. Liu N, Huang R, Baldacchino T, Sud A, Sud K, Khadra M, et al. Telehealth for Noncritical Patients With Chronic Diseases During the COVID-19 Pandemic. J Med Internet Res 2020 Aug 07;22(8):e19493 [FREE Full text] [doi: 10.2196/19493] [Medline: 32721925]
4. Guest JL, Del Rio C, Sanchez T. The Three Steps Needed to End the COVID-19 Pandemic: Bold Public Health Leadership, Rapid Innovations, and Courageous Political Will. JMIR Public Health Surveill 2020 Apr 06;6(2):e19043 [FREE Full text] [doi: 10.2196/19043] [Medline: 32240972]
5. Number of small and medium sized businesses in the healthcare industry in the United States in 2017, by NAICS category. Statista. URL: https://www.statista.com/statistics/785122/number-of-healthcare-smbs-by-naics-category-us, [accessed 2020-08-18]
6. Weiner R. Small medical practices struggle to survive amid coronavirus pandemic. The Washington Post. 2020 May 14. URL: https://www.washingtonpost.com/health/small-doctor-practices-struggle-to-survive-amid-coronavirus-pandemic/2020/05/14/328984e6-9390-11ea-91d7-cf4423d47683_story.html [accessed 2021-02-17]
7. National Health Expenditure Projections 2018-2027. Office of the Actuary at the Centers for Medicare and Medicaid Services. 2019. URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/ForecastSummary.pdf [accessed 2020-08-12]
8. Kaissi A, Charland T. The evolution of retail clinics in the United States, 2006-2012. Health Care Manag (Frederick) 2013;32(4):336-342. [doi: 10.1097/HCM.0b013e3182a9d73f] [Medline: 24168869]
9. Reid R, Ashwood J, Friedberg M, Weber E, Setodji C, Mehrotra A. Retail clinic visits and receipt of primary care. J Gen Intern Med 2013 Apr;28(4):504-512 [FREE Full text] [doi: 10.1007/s11606-012-2243-x] [Medline: 23070656]
10. Ambulatory Care Use and Physician office visits. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/fastats/physician-visits.htm [accessed 2020-08-12]
11. Abelson R. Doctors Without Patients: 'Our Waiting Rooms Are Like Ghost Towns'. The New York Times. 2020 May 05. URL: https://www.nytimes.com/2020/05/05/health/coronavirus-primary-care-doctor.html [accessed 2021-02-17]
12. Appold K. Independent Pharmacies: Not Dead Yet. Drug Topics: Voice of the Pharmacist. 2019 Jan 12. URL: https://www.drugtopics.com/view/independent-pharmacies-not-dead-yet [accessed 2021-02-17]
13. Van Nuys K, Joyce G, Ribero R, Goldman DP. Frequency and Magnitude of Co-payments Exceeding Prescription Drug Costs. JAMA 2018 Mar 13;319(10):1045-1047 [FREE Full text] [doi: 10.1001/jama.2018.0102] [Medline: 29536088]
14. Heath S. Gag Clause Bills Signed into Law, Improves Healthcare Transparency. Patient Engagement: Xtelligent Healthcare Media. 2018 Oct 11. URL: https://patientengagementhit.com/news/gag-clause-bills-signed-into-law-improves-healthcare-transparency [accessed 2020-08-12]
15. The Top 15 U.S. Pharmacies of 2019: Specialty Drugs Drive the Industry's Evolution. Drug Channels: Expert Insights on Pharmaceutical Economics and the Drug Distribution System. 2020 Mar 03. URL: https://www.drugchannels.net/2020/03/the-top-15-us-pharmacies-of-2019.html [accessed 2020-08-12]
16. Hopkins J. Mail-Order Drug Delivery Rises During Coronavirus Lockdowns. The Wall Street Journal. 2020 May 12. URL: https://www.wsj.com/articles/mail-order-drug-delivery-rises-during-coronavirus-lockdowns-11589281203 [accessed 2020-08-18]
17. Ientile G. NCPA Survey: Independent Pharmacies Report Financial Strain from COVID-19 Pandemic. Drug Topics: Voice of the Pharmacist. 2020 Apr 20. URL: https://www.drugtopics.com/view/ncpa-survey-independent-pharmacies-report-financial-strain-covid-19-pandemic [accessed 2020-08-12]
18. Jeremias S. Pharmacies Rise to the Challenges Imposed by COVID-19. AJMC: The Center for Biosimilars. 2020 May 13. URL: https://www.centerforbiosimilars.com/conferences/pqa-2020/pharmacies-rise-to-the-challenges-imposed-by-covid19 [accessed 2020-08-12]
19. Carter E. Pharmacists play a critical role in defeating COVID-19. The Hill. 2020 Jul 07. URL: https://thehill.com/blogs/congress-blog/healthcare/506136-pharmacists-play-a-critical-role-in-defeating-covid-19 [accessed 2020-08-12]
20. Jercich K. Senate HELP Committee weighs the future of telehealth. Healthcare IT News. 2020 Jun 17. URL: https://www.healthcareitnews.com/news/senate-help-committee-weighs-future-telehealth [accessed 2020-08-12]

Abbreviations

RVC: remote and virtual care
SCBS: satisfaction with the current business scenario
SRVC: satisfaction with remote and virtual care models
SUST: sustainability of the physician office, clinic, or pharmacy

Edited by C Basch, E Meinert; submitted 19.08.20; peer-reviewed by A Meyers, T Deng, M Dohan; comments to author 04.09.20; revised version received 09.09.20; accepted 31.01.21; published 25.02.21.

Please cite as:
Parthasarathy M, Khuntia J, Stacey R
Impact of Remote and Virtual Care Models on the Sustainability of Small Health Care Businesses: Perceptual Analysis of Small Clinics, Physician Offices, and Pharmacies in Colorado
J Med Internet Res 2021;23(2):e23658
URL: https://www.jmir.org/2021/2/e23658
doi:10.2196/23658
PMID:33539306

Original Paper

Health Care Workers' Reasons for Choosing Between Two Different COVID-19 Prophylaxis Trials in an Acute Pandemic Context: Single-Center Questionnaire Study

Alberto M Borobia[1*], MD, PhD; Irene García-García[1*], MD; Lucía Díaz-García[1], MD; Amelia Rodríguez-Mariblanca[1], MD; Lucía Martínez de Soto[1], MD; Jaime Monserrat Villatoro[1], MD; Enrique Seco Meseguer[1], MD; Juan J González[2], MD, PhD; Jesús Frías Iniesta[1], MD, PhD; Elena Ramírez García[1], MD, PhD; Jose Ramón Arribas[2], MD, PhD; Antonio J Carcas-Sansuán[1], MD, PhD

[1]Clinical Pharmacology Department, La Paz University Hospital (IdiPAZ), School of Medicine, Universidad Autónoma de Madrid, Madrid, Spain
[2]Internal Medicine Department, La Paz University Hospital (IdiPAZ), School of Medicine, Universidad Autónoma de Madrid, Madrid, Spain
[*]these authors contributed equally

Corresponding Author:
Alberto M Borobia, MD, PhD
Clinical Pharmacology Department, La Paz University Hospital (IdiPAZ)
School of Medicine, Universidad Autónoma de Madrid
Paseo Castellana, 261
Madrid, 28046
Spain
Phone: 34 912071466
Fax: 34 912071466
Email: alberto.borobia@idipaz.es

Abstract

Background: In April 2020, two independent clinical trials to assess SARS-CoV-2 prophylaxis strategies among health care workers were initiated at our hospital: MeCOVID (melatonin vs placebo) and EPICOS (tenofovir disoproxil/emtricitabine vs hydroxychloroquine vs combination therapy vs placebo).

Objective: This study aimed to evaluate the reasons why health care workers chose to participate in the MeCOVID and EPICOS trials, as well as why they chose one over the other.

Methods: Both trials were offered to health care workers through an internal news bulletin. After an initial screening visit, all subjects were asked to respond to a web-based survey.

Results: In the first month, 206 health care workers were screened and 160 were randomized. The survey participation was high at 73.3%. Health care workers cited "to contribute to scientific knowledge" (n=80, 53.0%), followed by "to avoid SARS-CoV-2 infection" (n=33, 21.9%) and "the interest to be tested for SARS-CoV-2" (n=28, 18.5%), as their primary reasons to participate in the trials. We observed significant differences in the expected personal benefits across physicians and nurses (P=.01). The vast majority of volunteers (n=202, 98.0%) selected the MeCOVID trial, their primary reason being their concern regarding adverse reactions to treatments in the EPICOS trial (n=102, 69.4%).

Conclusions: Health care workers' reasons to participate in prophylaxis trials in an acute pandemic context appear to be driven largely by their desire to contribute to science and to gain health benefits. Safety outweighed efficacy when choosing between the two clinical trials.

(J Med Internet Res 2021;23(2):e23441) doi:10.2196/23441

KEYWORDS

clinical trials; COVID-19; health care worker; motivation; personnel; pre-exposure; professional practice; prophylaxis; SARS-CoV-2; volunteers, web-based survey; workplace safety

Introduction

Therapeutic and phase I clinical studies have primarily focused on motivations for participating in clinical trials. However, limited information is available regarding subjects' motivators to participate in prophylaxis trials in an acute pandemic context, with particular reference to COVID-19, especially in Western countries. In this setting, when several trials compete in recruiting participants from the same center, the participants' reasons for choosing one trial over the other can be very relevant.

Currently, no pre-exposure prophylaxis therapy has been approved for COVID-19, but numerous clinical trials have been initiated in Europe and the United States, most of them assessing the potential of hydroxychloroquine as a prophylactic agent (eg, EudraCT-2020-001565-37, EudraCT-2020-001536-98, NCT04352946, NCT04354870, and NCT04328467) [1-5].

As of April 21, 2020, two independent clinical trials to assess COVID-19 prophylaxis strategies among health care workers have been initiated at the Clinical Trials Unit, La Paz University Hospital (Madrid, Spain): the MeCOVID trial (melatonin vs placebo; EudraCT-2020-001530-35) and the EPICOS trial (a four-arm clinical trial comparing the efficacy of tenofovir disoproxil/emtricitabine, hydroxychloroquine, combination therapy, and placebo; EudraCT-2020-001385-11).

This study aimed to evaluate the motivations among health care workers at our hospital to participate in COVID-19 prophylaxis trials and the reasons to select one of the two trials currently underway at our hospital.

Methods

A choice between the two clinical trials was offered to our hospital personnel through the internal news bulletin. All potential participants received the information regarding both clinical trials and had the opportunity to have any questions answered before choosing between the two studies. Their choices were not influenced by the team of investigators.

Participants were screened to their chosen trial. In both clinical trials, they underwent COVID-19 screening using a serologic rapid test (Orient Gene, Orient Gene Biotech; or Wondfo, Guangzhou Wondfo Biotech). All screened subjects were asked to complete a web-based survey after their screening visit.

The questionnaire consisted of 9 questions on participant demographics, professional designation and work site, knowledge of COVID-19, motivation to participate, reasons to choose one trial over the other, and treatment expectations (Multimedia Appendix 1).

The survey was administered as a Google form. An invitation email containing a link to the questionnaire was distributed among all screened volunteers. No reminder emails were sent to nonresponders. Participants' responses were anonymized and maintained confidential in accordance with Google's privacy policy.

Descriptive statistics including variables with percentage values were determined. We performed the Shapiro–Wilk test to determine whether age was normally distributed in our study population; accordingly, we rejected our null hypothesis in the test for normality ($P<.001$) and found that age was normally distributed.

Statistical analyses were performed using R software (version 3.6.3, The R Foundation). We analyzed differences between age and position using the Mann–Whitney *U* test and between sex and position using the chi-square test. Furthermore, we performed the Fisher exact test to analyze differences in the responses of physicians and nurses.

Results

In the first month of recruitment, 206 health care workers were screened and 160 were randomized (MeCOVID: n=156, 97.5% and EPICOS: n=4, 2.5%). Volunteers selected the trial in which they wanted to participate, the vast majority (n=202, 98.0%) having selected the MeCOVID trial. Table 1 shows the distribution of survey outcomes in the sampling frame. Furthermore, Table 2 shows the main data and the survey findings stratified by designation (physicians vs nurses vs others).

Table 1. Distribution of outcomes in the sampling frame (N=206).

Questionnaire response status among health care workers	Participants, n (%)
Returned the questionnaire	151 (73.3)
Complete	142 (68.9)
Partial or break-off with sufficient information[a]	9 (4.4)
Did not return the questionnaire (nonresponders)	55 (26.7)
Logged in to survey but did not complete any items[b]	16 (7.8)
Returned responses[c]	29 (14.1)
Invitation returned undelivered[d]	10 (4.8)

[a]Registered for the survey and responded to almost all questions (maximum of 2 questions not answered).

[b]Registered for the survey but returned no responses.

[c]Remaining sample (participants remaining after excluding those who returned the questionnaires, those who logged in to survey but did not complete any items, and those for whom the invitation was returned undelivered.

[d]Email invitation was returned undelivered. Email delivery failed owing to the use of an incorrect, outdated, or out-of-space email address.

Table 2. Main survey findings stratified by participant designation (N=151).

Survey findings	Participant designation			*P* value[a]
	Physicians (n=64)	Nurses[b] (n=79)	Others[c] (n=8)	
Responder characteristics				
Sex (males), n (%)	27 (42.2)	9 (11.4)	1 (12.5)	.001
Age (years), median (IQR)	41 (31-51)	37 (29-46)	46 (31-56)	.14
Knowledge of COVID-19, n (%)				.001
Expert	39 (60.9)	16 (20.2)	1 (12.5)	
Basic	25 (39.1)	55 (69.6)	7 (87.5)	
Some	0 (0)	7 (8.9)	0 (0)	
No knowledge	0 (0)	1 (1.3)	0 (0)	
Site of work, n (%)				.01
Emergency department	8 (12.5)	16 (20.3)	6 (75.0)	
Intensive care units	4 (6.3)	18 (22.8)	1 (12.5)	
Hospitalization wards	31 (48.4)	26 (32.9)	1 (12.5)	
External offices/other	21 (32.8)	19 (24.0)	0 (0)	
Main motivators, n (%)				.03
Scientific knowledge	31 (48.5)	42 (53.2)	7 (87.5)	
To prevent SARS-CoV-2 infection	13 (20.3)	20 (25.3)	0 (0)	
To have access to a SARS-CoV-2 rapid test	18 (28.1)	9 (11.4)	1 (12.5)	
Other	2 (3.1)	8 (10.1)	0 (0)	
MeCOVID trial participants, n	62	77	8	
Reason to participate in the MeCOVID trial (multiple answer question), n (%)				<.001
Adverse reactions to drugs in the EPICOS trial	53 (85.5)	49 (63.6)	0 (0)	
Melatonin being a sleep aid	6 (9.7)	17 (22.1)	1 (12.5)	
Melatonin having a high efficacy	2 (3.2)	17 (22.1)	6 (75.0)	
Contraindications regarding EPICOS drugs	1 (1.6)	4 (5.2)	0 (0)	
Not reported	4 (6.5)	1 (1.3)	1 (12.5)	
Expectations regarding the MeCOVID trial, n (%)				.13
Drug efficacy and safety	26 (42.0)	50 (64.9)	8 (100)	
Drug efficacy with adverse reactions	2 (3.2)	14 (18.2)	0 (0)	
No drug efficacy and no adverse reactions	31 (50.0)	13 (16.9)	0 (0)	
No drug efficacy with adverse reactions	1 (1.6)	0 (0)	0 (0)	
Not reported	2 (3.2)	0 (0)	0 (0)	

[a]*P* values refer to differences between the responses of physicians and nurses. “Others” have been excluded from the analysis owing to their small sample size.

[b]“Nurses” includes nurse practitioners (n=54) and nursing assistants (n=25).

[c]“Others” includes laboratory technicians (n=2) and ancillary personnel (n=6).

In total, 56 (37.1%) participants reported having expert knowledge of COVID-19, of whom 39 (69.6%) were physicians, 16 (28.6%) were nurses, and 1 (1.8%) had another designation. The main motivation for many to participate in the trials was “to contribute to scientific knowledge” (n=80, 53.0%), followed by “to avoid SARS-CoV-2 infection” (n=33, 21.9%) and “the interest to be tested for SARS-CoV-2 by a serologic rapid test” (n=28, 18.5%) (Multimedia Appendix 2).

Regarding the expected personal benefits, we observed significant differences (*P*=.01) in the responses of physicians and nurses: the main expected benefit among physicians (n=18, 54.6%) was to gain access to SARS-CoV-2 rapid tests, while that among nurses (n=20, 62.5%) was to prevent SARS-CoV-2 infection (Multimedia Appendix 3).

Among the 147 subjects in the MeCOVID trial, the main motivation to participate was "the fear to present any toxicity related to EPICOS treatments" (n=102, 69.4%), while that among participants in the EPICOS trial was "the belief that hydroxychloroquine and tenofovir/emtricitabine might be more effective than melatonin to prevent SARS-CoV-2 infection" (n=2, 50%).

Again, we observed significant differences in the responses of physicians and nurses with respect to their reasons for selecting the MeCOVID trial. In total, 53 (85.5%) physicians and 49 (63.6%) nurses reported choosing the MeCOVID trial owing to adverse reactions to drugs used in the EPICOS trial. Conversely, 17 (22.1%) nurses and only 2 (3.2%) physicians preferred the MeCOVID trial to the EPICOS trial owing to the higher efficacy of melatonin, and 17 (22.1%) nurses and only 6 (9.7%) physicians reported the use of melatonin as a sleep aid (P<.001).

We had expected most participants (n=84, 57.1%) to favor the MeCOVID trial primarily because melatonin would be efficacious and safe, with 44 (29.9%) participants believing that although melatonin would not be efficacious, it would at least not lead to adverse reactions. Most participants (n=3, 75.0%) favoring the EPICOS trial expected the treatment to not only be efficacious but also to lead to some adverse reactions.

Discussion

Principal Findings

Prevention of SARS-CoV-2 transmission among health care workers is key to managing COVID-19 outbreaks. Although the precise number of COVID-19–infected health care workers is lacking, and studies have reported this infection rate to be 20% in Spain, approximately 10% in Italy and the United States, 6% in the Netherlands, and 3.8% in China [6]. Drug prophylaxis is one of the measures proposed to prevent SARS-CoV-2 infection among health care workers, and it is a priority to obtain strong evidence for this intervention.

This study evaluated the motivations among health care workers at our hospital to participate in 1 of 2 COVID-19 prophylaxis trials—MeCOVID and EPICOS—and their reasons to choose one trial over the other, given the different characteristics of these trials. Survey participation was high (n=151, 73.3%), with a response rate of approximately 70%. In our opinion, the reason for this is the use of a web-based survey, which was easy to respond to (taking only approximately 5 minutes to complete).

Empirical studies have reported that altruism and self-interest are the two primary motivations for participants to enroll in unpaid clinical trials. This holds good in various clinical situations, including participation in prophylaxis and vaccines trials, and in trials conducted at different geographical locations [7-10].

To our knowledge, this study is the first of its kind to be conducted in an acute pandemic context. Herein, the motivations of health care workers, who are aware of the severity and complications of COVID-19, to participate in the trials were similar to those reported in other clinical situations; that is, to contribute to science (n=80, 53%) and to gain personal benefits (n=71, 47.0%), and no significant differences were observed in the responses of physicians and nurses. However, the expected personal benefits were significantly different between physicians and nurses, with physicians being more interested in gaining access to SARS-CoV-2 rapid tests and nurses expecting to prevent SARS-CoV-2 infection. Such a combination of motivators has also been reported among subjects participating in HIV prophylaxis trials and Ebola vaccine trials [11-14].

A major reason for participants to choose one trial over the other in this study was the fear of adverse reactions to drugs used in the EPICOS trial; hence, two-thirds of the volunteers opted to participate in the MeCOVID trial. Our volunteers seemed to participate in the MeCOVID trial because they perceived it as less risky (they opted for this trial despite one-third volunteers not expecting any treatment efficacy). Furthermore, the reasons to participate in the MeCOVID trial were significantly different between physicians and nurses, with physicians being more concerned about adverse drug reactions.

Strengths and Limitations

Although our study has a large sample and a high survey participation rate, it has some limitations of note. There are some biases in our study: a selection bias owing to the single-center study design and a low participation rate in the EPICOS trial (n=4, 2.5%). However, one of our study objectives was to evaluate motivations of health care workers to participate in COVID-19 prophylaxis trials independent of the trial selected. A large number of physicians and nurses participated in our survey, which enabled us to compare the responses between them. Another study limitation is that we could not precisely measure the reception of the survey via email. To be conservative and to estimate the survey participation rate, we have included the number of emailed individuals in the denominator.

Conclusions

Our results show that health care workers' motivations to participate in prophylaxis trials in an acute pandemic context appear to be driven mostly by their desire to contribute to science and obtain some health benefits. When the opportunity to participate in various trials is offered, safety markedly outweighed efficacy among the participants. Most participants opted to participate in one study over the other for safety reasons, having selected the trial they perceived as less risky, even if they consider it less efficacious.

Authors' Contributions

AMB, AJC, JRA, and IGG conceptualized the study. IGG, JF, and ER designed the study. AMB, AJC, and IGG analyzed the data. LDG, ARM, LMS, JMV, ESM, and JJG performed the investigation. IGG curated the data. AMB and IGG drafted the

manuscript. AJC and JRA reviewed and edited the manuscript. All authors have read and approved the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey for volunteers participating in clinical trials for COVID-19 porphylaxis.
[PDF File (Adobe PDF File), 83 KB - jmir_v23i2e23441_app1.pdf]

Multimedia Appendix 2
Main motivations to participate in the trials classified as to contribute to science and personal benefits, subgrouped by physician vs nurses.
[DOC File , 42 KB - jmir_v23i2e23441_app2.doc]

Multimedia Appendix 3
Personal benefits to participate in the trials subgrouped by physician vs nurses.
[DOC File , 34 KB - jmir_v23i2e23441_app3.doc]

References

1. Pre-exposure prophylaxis with hydroxychloroquine for high-risk healthcare workers during the COVID-19 pandemic (prep_covid): a unicentric, double-blinded randomized controlled trial. EU Clinical Trials Register. 2020. URL: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2020-001565-37 [accessed 2020-11-04]
2. Prophylaxis of COVID-19 infection with hydroxychloroquine in healthcare personnel with high risk of infection. EU Clinical Trials Register. 2020. URL: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2020-001536-98 [accessed 2020-11-04]
3. HEalth Care Worker pROphylaxis Against COVID-19: The HERO Trial (HERO). ClinicalTrials.gov. 2020. URL: https://clinicaltrials.gov/ct2/show/NCT04352946?term=NCT04352946&draw=2&rank=1 [accessed 2020-11-04]
4. COVID-19 PrEP HCW HCQ Study. ClinicalTrials.gov. 2020. URL: https://clinicaltrials.gov/ct2/show/NCT04354870?term=NCT04354870&draw=2&rank=1 [accessed 2020-11-04]
5. Pre-exposure Prophylaxis for SARS-Coronavirus-2. ClinicalTrials.gov. 2020. URL: https://clinicaltrials.gov/ct2/show/NCT04328467?term=NCT04328467&draw=2&rank=1 [accessed 2020-11-04]
6. Coronavirus disease 2019 (COVID-19) in the EU/EEA and the UK – ninth update. European Centre for Disease Prevention and Control. 2020 Apr 23. URL: https://www.ecdc.europa.eu/sites/default/files/documents/covid-19-rapid-risk-assessment-coronavirus-disease-2019-ninth-update-23-april-2020.pdf [accessed 2020-05-01]
7. Olsen L, DePalma L, Evans JH. Self-Interested and Altruistic Motivations in Volunteering for Clinical Trials: A More Complex Relationship. J Empir Res Hum Res Ethics 2020 Dec;15(5):443-451. [doi: 10.1177/1556264620914463] [Medline: 32363984]
8. Browne JL, Rees CO, van Delden JJM, Agyepong I, Grobbee DE, Edwin A, et al. The willingness to participate in biomedical research involving human beings in low- and middle-income countries: a systematic review. Trop Med Int Health 2019 Mar;24(3):264-279 [FREE Full text] [doi: 10.1111/tmi.13195] [Medline: 30565381]
9. Moorcraft SY, Marriott C, Peckitt C, Cunningham D, Chau I, Starling N, et al. Patients' willingness to participate in clinical trials and their views on aspects of cancer research: results of a prospective patient survey. Trials 2016 Jan 09;17:17 [FREE Full text] [doi: 10.1186/s13063-015-1105-3] [Medline: 26745891]
10. Gaul C, Malcherczyk A, Schmidt T, Helm J, Haerting J. Motivation of patients to participate in clinical trials. An explorative survey. Med Klin (Munich) 2010 Feb;105(2):73-79. [doi: 10.1007/s00063-010-1016-4] [Medline: 20174906]
11. Nyaoke B, Museve E, Masheti M. Motivators to Participate in a Non-HIV Experimental Study among HIV-Infected Individuals in Nairobi, Kenya: A Consecutive Sample of an Ebola Vaccine Clinical Trial. J Vaccines Immunol: JVII-118 2017 Nov 27:10 [FREE Full text] [doi: 10.29011/2575-789x.000118]
12. Nyaoke BA, Mutua GN, Sajabi R, Nyasani D, Mureithi MW, Anzala OA. Volunteer motivators for participating in HIV vaccine clinical trials in Nairobi, Kenya. PLoS One 2017;12(9):e0183788 [FREE Full text] [doi: 10.1371/journal.pone.0183788] [Medline: 28880880]
13. Cattapan A, Browne K, Halperin DM, Di Castri A, Fullsack P, Graham J, et al. Motivation for participating in phase 1 vaccine trials: Comparison of an influenza and an Ebola randomized controlled trial. Vaccine 2019 Jan 07;37(2):289-295. [doi: 10.1016/j.vaccine.2018.11.014] [Medline: 30528592]
14. Fisher CB, Fried AL, Ibrahim Puri L, Macapagal K, Mustanski B. "Free Testing and PrEP without Outing Myself to Parents:" Motivation to participate in oral and injectable PrEP clinical trials among adolescent men who have sex with men. PLoS One 2018;13(7):e0200560 [FREE Full text] [doi: 10.1371/journal.pone.0200560] [Medline: 30044845]

Edited by G Fagherazzi; submitted 12.08.20; peer-reviewed by C García, P Kanzow; comments to author 03.11.20; revised version received 11.11.20; accepted 01.02.21; published 25.02.21.

Please cite as:
Borobia AM, García-García I, Díaz-García L, Rodríguez-Mariblanca A, Martínez de Soto L, Monserrat Villatoro J, Seco Meseguer E, González JJ, Frías Iniesta J, Ramírez García E, Arribas JR, Carcas-Sansuán AJ
Health Care Workers' Reasons for Choosing Between Two Different COVID-19 Prophylaxis Trials in an Acute Pandemic Context: Single-Center Questionnaire Study
J Med Internet Res 2021;23(2):e23441
URL: https://www.jmir.org/2021/2/e23441
doi:10.2196/23441
PMID:33556032

Original Paper

Effects of COVID-19 Emergency Alert Text Messages on Practicing Preventive Behaviors: Cross-sectional Web-Based Survey in South Korea

Minjung Lee[1,2], PhD; Myoungsoon You[1], MPH, PhD

[1]Department of Public Health Sciences, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea

[2]Office of Dental Education, School of Dentistry, Seoul National University, Seoul, Republic of Korea

Corresponding Author:
Myoungsoon You, MPH, PhD
Department of Public Health Sciences, Graduate School of Public Health
Seoul National University
Gwanak-ro 1, Gwanak-gu
Seoul, 08826
Republic of Korea
Phone: 82 880 2773
Email: msyou@snu.ac.kr

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2021/3/e28660

Abstract

Background: Sending emergency messages via mobile phone text messaging can be a promising communication tool to rapidly disseminate information and promote preventive behavior among the public during epidemic outbreaks. The battle to overcome COVID-19 is not yet over; thus, it is essential that the public practices preventive measures to prevent the spread of COVID-19.

Objective: This study aimed to investigate the effectiveness of reading and obtaining information via emergency alert SMS text messages and their effects on the individual's practice of preventive behaviors during the early stages of the COVID-19 outbreak in South Korea.

Methods: A cross-sectional web-based survey comprising 990 participants was conducted over 3 days (March 25-27, 2020). A multivariable logistic regression analysis revealed the sociodemographic factors that might influence the behavior of reading emergency alert text messages. A hierarchical linear regression model estimated the associations between reading emergency alert text messages for each precautionary behavior practiced against COVID-19. Additionally, the indirect effects of reading the text messages on each precautionary behavior via psychological factors (ie, perceived risk and response efficacy) were calculated. All data were weighted according to the 2019 Korea census data.

Results: Overall, 49.2% (487/990) of the participants reported that they always read emergency alert text messages and visited the linked website to obtain more information. Factors such as female sex (odds ratio [OR] 1.68, 95% CI 1.28-2.21) and older age (30-39 years: OR 2.02, 95% CI 1.25-3.28; 40-49 years: OR 2.84, 95% CI 1.80-4.47; 50-59 years: OR 3.19, 95% CI 2.01-5.06; 60 years and above: OR 3.12, 95% CI 2.00-4.86 versus 18-29 years) were identified to be associated with a higher frequency of reading the text messages. Participants who always read the text messages practiced wearing facial masks (β=.074, P=.01) more frequently than those who did not. In terms of social distancing, participants who reported they always read the text messages avoided crowded places (β=.078, P=.01) and canceled or postponed social gatherings (β=.103, P<.001) more frequently than those who did not read the text messages. Furthermore, reading text messages directly and indirectly affected practicing precautionary behaviors, as the mediation effect of response efficacy between reading text messages and practicing preventive behaviors was significant.

Conclusions: Our findings suggest that emergency alert text messages sent to individuals' mobile phones are timely and effective strategies for encouraging preventive behavior in public. Sending emergency alert text messages to provide the public with accurate and reliable information could be positively considered by the health authorities, which might reduce the negative impact of infodemics.

(J Med Internet Res 2021;23(2):e24165) doi:10.2196/24165

KEYWORDS

COVID-19; coronavirus; preventive behaviors; text message; mobile phone; alert; prevention; behavior; public health; survey

Introduction

In recent years, mobile technology and text messages have emerged as a promising communication tool to rapidly disseminate information during several emergencies [1-3]. There is a widespread use of and access to smartphone and mobile devices, with mobile phone technology penetration at nearly 100% worldwide. South Korea has among the highest ownership percentages (94%) [4]. Their improved geolocation capacity and access to broadband and satellite communication infrastructure have enabled emergency alert SMS text messages to be sent to end users directly [5]. One of the attractive features of mobile text communication for emergency communication is the ability to target text messages to all phones in a specific location very quickly [6,7], making such communication highly efficient. Moreover, unlike some other media sources, readers can read, reread, and analyze the information provided via text messages [8]. A previous study in the United Kingdom suggests that a system that sends emergency messages via mobile phone text messaging would be generally well accepted by the public and likely to improve uptake of protective behaviors when combined with other approaches [9].

After the outbreak of Middle East Respiratory Syndrome (MERS) in 2015, South Korea passed the Framework Act on the Management of Disasters and Safety. This Act established disaster and safety management systems for state and local governments and prescribed matters necessary for disaster management to protect citizens' lives, bodies, and properties. The Act included new regulations for emergency management and response to improve safety, such as requiring owners or managers of telecommunication facilities' preferential capacity to forecast, alert, notify, or undertake other emergency measures concerning a disaster when necessary. The central and local governments of South Korea have sent emergency alert text messages to the public during the COVID-19 outbreak. In general, emergency alert text messages include two types of content. In the first type, the central government sends persuasive messages, encouraging individuals to take preventive measures. In the second type, local governments also recommend precautionary behaviors but mostly send risk information such as the number of confirmed cases in the residence area, contact tracing of confirmed cases, and closure of certain places. Examples of text messages sent to the public in March 2020 are shown in Table 1.

Table 1. Examples of emergency alert text messages sent to the public during the COVID-19 pandemic.

Sender and message contents (translated)	Date sent
Central governments (Central Disaster and Safety Countermeasures Headquarters of Korea)	
Join us for a"two-week stop"to protect your family and friends! Thank you for your dedication. [10]	March 23, 2020
Observe strict hygiene rules when caring for the elderly.	March 24, 2020
Even if you are apart, your heart is close. [11]	
Local governments	
Three additional COVID-19 confirmedcases occurred. An epidemiological investigation is currently underway, and the route of confirmed persons is scheduled to be released on the website after confirmation. (Seocho City Office)	March 13, 2020
Check the route of the 5th COVID-19 confirmed case of Jeollanam-do (Hwasun) on the county office's website, and if the route overlaps, please consult with Jangseong Community Health Center (061-360-8333).	March 17, 2020

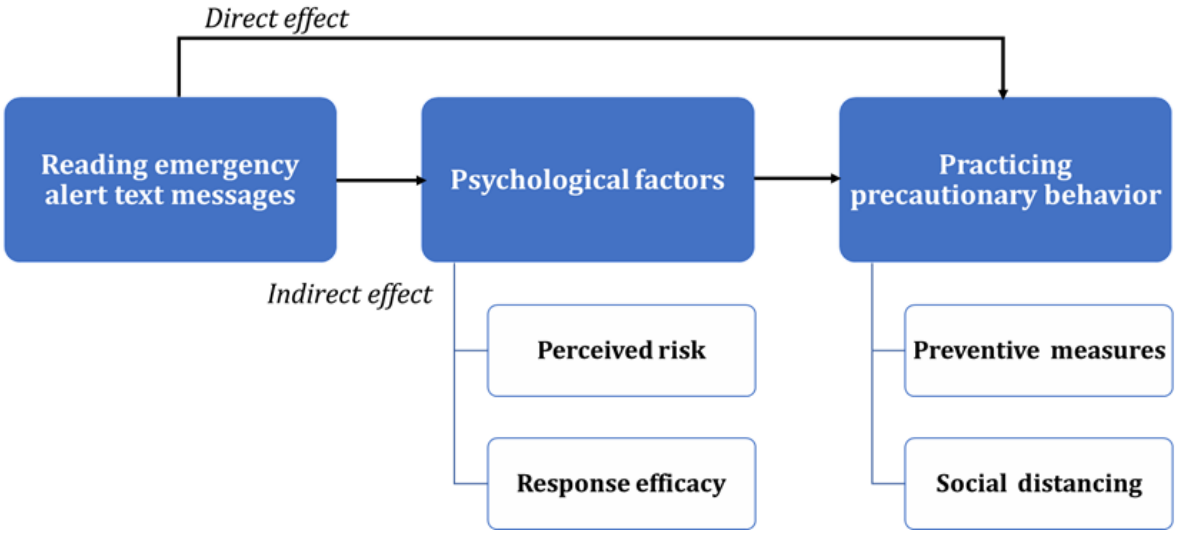
The battle to overcome COVID-19 is not yet over. Therefore, nonpharmaceutical interventions such as wearing facial masks and practicing social distancing are critical. Efforts to sustain and elevate these practices by the public are among the most important goals of public health authorities. Effective public health risk communication about the outbreak and guidance on how to respond can alleviate the negative impacts of the public health emergency and save lives [12]. There can be negative consequences if the public does not practice preventive behaviors quickly enough [13]. Thus, public health education and public health policies should be implemented so that "social learning," which may be described as the collective effects of communication efforts, can take place since the public is one of the most critical stakeholders in combatting the outbreak [13,14].

Successful emergency communication is determined by how quickly or reliably a message can be disseminated and how people respond to the information they receive [15,16]. Emergency alert messages have no benefit unless the readers read the message and follow the provided guidelines. Therefore, understanding why people accept or resist responding to warning and alert messages is essential. Some researchers have proposed decision models that could harmoniously explain recipient behaviors in response to emergency alerts [17-20]. These models include protective action decision-making [18], protection motivation theory [20], person-relative-to-event theory [21], and the theory of planned behavior [17]. According to the protection motivation theory, threat appraisal depends on individual perceptions of disease severity (ie, evaluating the state of the environment and observing what happens to others) and the individual's susceptibility. A coping appraisal is driven

by the efficacy belief, which comprises perceived response efficacy (ie, the belief that the recommended behavior will protect) and self-efficacy (ie, the ability to perform the recommended behavior). We examined two psychological concepts by adopting the protection motivation theory to explore preventive behaviors among individuals—perceived risk and response efficacy.

In this study, we investigated the effectiveness of reading and obtaining information from emergency alert text messages and their effects on the individual's preventive behaviors during the COVID-19 outbreak in South Korea. First, we examined the extent to which people read the text messages and identified sociodemographic factors that contribute to this behavior. Second, we determined whether the individuals responded to these messages by investigating the direct effect of reading text messages on practicing preventive behaviors. Finally, we examined whether the effect of reading text messages on practicing precautionary behaviors was mediated via psychological factors by investigating indirect effects through perceived risk of COVID-19 infection and response efficacy of each behavior. The framework of this study was constructed as shown in Figure 1. To our knowledge, most previous studies evaluating the effect of emergency alert text messages depended on behavioral intentions, mainly in hypothetical emergencies [9,22-24]. These previous studies are limited because behavioral intent in hypothetical situations may not accurately predict behavioral responses in actual situations [25,26].

Figure 1. Framework used in this study.



Methods

Study Design and Sampling

A cross-sectional web-based survey design was adopted to (1) evaluate preventive behaviors of the public and (2) assess whether they read emergency alert text messages during the COVID-19 pandemic, by using an anonymous questionnaire. The survey was conducted via a web-based platform developed by a research company called Korea Research. The company recruited respondents by sending survey invitations containing general information about the survey, such as its aim and consent statement via email or text messages, to registered survey panel members who met the inclusion criteria. The inclusion criteria were as follows: (1) aged 18 years or older, (2) a resident of South Korea, and (3) a Korean speaker.

The recruiting company sampled respondents using a proportional quota sampling process based on age, sex, and geographic region. The respondents were asked to provide electronic informed consent on the first page of the survey. Korea Research is responsible for protecting the confidentiality of the anonymous survey respondents. Over 1000 participants completed the surveys, of which 990 were included in the analysis after excluding incomplete responses. The participation rate was 62.5%, which is assumed to be acceptable for web-based surveys [27-29]. The present study protocol was reviewed and approved by the Institutional Review Board at Seoul National University (IRB No. 2003/002-005), Seoul, South Korea. All participants, upon enrollment, provided their informed consent. The data collection took place over 3 days (March 25-27, 2020), about 2 months after the Korea Centers for Disease Control and Prevention confirmed the first case of COVID-19 during the early stage of the pandemic and just before 10,000 confirmed cases had been reported (ie, on April 3, 2020).

Measurements

Dependent Variables

Outcome variables were precautionary behaviors related to the threat of COVID-19. They were classified into one of two categories: (1) preventive measures (eg, wearing facial masks and practicing hand hygiene) and (2) social distancing behaviors (eg, avoiding crowded places and postponing or canceling social events). Participants self-reported the frequency of the actions they undertook during the previous week by using a 4-point Likert-type scale (1=never, 2=sometimes, 3=often, and 4=always).

Independent Variables

Whether the participants read COVID-19 emergency alert text messages was assessed by the following question: “Do you read the COVID-19 emergency alert text messages received?”

Participants responded by choosing one of four options: “I never read the messages” (score=1); “I rarely read the messages” (score=2); “I read the messages occasionally and visit the linked website if necessary” (score=3); or “I always read the message and visit the linked site to get more information” (score=4). Participants who reported they never or rarely read these text messages were also asked why. They were then grouped into binary groups to reduce the likelihood of low content validity due to the survey item; these groups included two behaviors: reading the message and visiting the linked website. The binary groups consisted of (1) participants who always read the text messages and visited the linked website to get more information (score=4) and (2) those who do not (score=1 for “always” and score=0 for “otherwise”).

Among the psychological factors, perceived risk of COVID-19 infection comprised (1) perceived susceptibility, signifying an individual's beliefs about their possibility of infection, and (2) perceived severity, signifying the seriousness of infection [30]. Participants were asked, “What do you think is the possibility of you contracting COVID-19 infection?” and “What do you think will be the severity if you are infected with COVID-19?” Responses were rated on a 5-point Likert-type scale, with responses ranging from “very low” (score=1), “neither low nor high” (score=3), to “very high” (score=5). To promote response efficacy, participants answered the question, “To what extent do you think the precautionary behavior is an effective way to reduce the risk of COVID-19 infection?” for each behavior examined in this study [31]. Their responses were rated on a 4-point Likert-type scale, ranging from “not at all” (score=1) to “extremely” (score=4).

Sociodemographic factors included sex (score=1 for male, score=2 for female), age, and the presence of children younger than elementary school age at home (score=1 for more than 1 child, score=0 for no children). We also assessed the participants’ education level (score=1 for middle school or below, score=2 for high school graduate, score=3 for college and above) and monthly household income reported in their local currency (ie, KRW: score=1 for below KRW 2 million [US $1800], score=2 for KRW 2-3.99 million [US $1800-3600], score=3 for KRW 4-5.99 million [US $3600-5400], and score=4 for KRW 6 million [US $5400] or above). We collected information about the participants’ occupation and whether they were employed, as well as the presence of underlying disease. We also asked the participants to report any diagnoses for the underlying diseases (eg, hypertension, dyslipidemia, diabetes, chronic cardiac disease, asthma, and cancer). We then grouped the participants based on with or without a diagnosis for one or more underlying diseases.

Covariates

Other measures associated with the exposure to COVID-19–related alert text messages in other information channels served as covariates in the hierarchical linear regression analysis. Media exposure was measured by the following question: “In the past week, how often did you use media every day to learn about the news?” with answers ranging from “never” (score=1) to “quite often” (score=4). Additionally, we measured the variable of obtaining COVID-19–related information. Participants rated the following questions on a 4-point Likert scale: “Did you actively search for information about COVID-19 during the last week?” with responses ranging from “never” (score=1) to “very often” (score=4).

Statistical Analysis

All data were weighted by age, sex, and geographic region distributions in South Korea, according to the Korean Statistical Information Service (2019) [32]. The sampling weights were provided by Korea Research, the company that conducted the survey, by using a random iterative method weighting process, in which each participant was assigned a single weight value. Sampling weights were incorporated into all analyses, and corresponding analytic methods were used. Sampling weights were based on all cases from the present study. Statistical analyses were conducted using R version 4.0.2 software (R Foundation for Statistical Computing). All quantitative variables are reported as absolute numbers with proportions (%) and mean values with SD.

The two groups for reading text messages included (1) those who always read and visited the linked website for more information and (2) those who did not. We measured group differences according to participants’ sociodemographic characteristics, psychological factors, and preventive behaviors, which were analyzed by a chi-square test or *t* test, as appropriate (Table 2). A multivariable logistic regression revealed which sociodemographic factors (eg, sex, age, educational level, monthly household income, employment status, presence of young children in the household, and presence of underlying disease) might influence emergency alert text message reading behavior.

Table 2. Characteristics of survey participants.

Variables	Participants (N=990)	Weighted[a] values (N=990)	Participants who read text messages		*P* value[b,c]
			Always (n=487)	Otherwise (n=503)	
Sociodemographic factors					
Sex, n (%)					
Male	475 (47.9)	491 (49.6)	209 (42.8)	282 (56.1)	<.001[b]
Female	515 (52)	499 (50.4)	278 (57.2)	221 (43.9)	
Age (years), mean (SD)	47.07 (14.8)	46.45 (15.1)	48.89 (14.0)	44.08 (15.6)	<.001[c]
18-29, n (%)	159 (16.1)	176 (17.8)	52 (10.7)	124 (24.7)	<.001[b]
30-39, n (%)	157 (15.9)	160 (16.2)	73 (15)	87 (17.3)	
40-49, n (%)	197 (19.9)	192 (19.4)	103 (21.2)	89 (17.7)	
50-59, n (%)	205 (20.7)	197 (19.9)	109 (22.4)	88 (17.5)	
≥60, n (%)	272 (27.5)	264 (26.7)	150 (30.7)	115 (22.9)	
Education level, n (%)					
Middle school or below	28 (2.8)	29 (2.9)	11 (2.3)	18 (3.6)	.46[b]
High school graduate	474 (47.9)	475 (48)	237 (48. 7)	238 (47.3)	
College and above	488 (49.3)	486 (49.1)	239 (49.1)	247 (49.1)	
Monthly household income[d] (million KRW), n (%)					
<2	127 (12.8)	127 (12.9)	67 (13.8)	61 (12.1)	.27[b]
2-3.99	312 (31.5)	313 (31.6)	156 (32.1)	157 (31.2)	
4-5.99	260 (26.3)	259 (26.2)	134 (27.6)	124 (24.7)	
≥6	291 (29.4)	290 (29.3)	130 (26.5)	161 (32)	
Occupation status, n (%)					
Out of labor	387 (39.1)	387 (39.1)	287 (59.1)	314 (62.4)	.29[b]
Working	603 (60.9)	602 (60.9)	200 (40.9)	189 (37.6)	
Presence of children, n (%)					
None	894 (90.3)	894 (90.3)	441 (90.6)	453 (90.1)	.79[b]
More than 1	96 (9.7)	96 (9.7)	46 (9.5)	50 (9.9)	
Underlying disease, n (%)					
None	583 (58.9)	590 (59.7)	277 (57)	313 (62.2)	.09[b]
More than 1	407 (41.1)	399 (40.4)	210 (43)	190 (37.8)	
Psychological factors					
Perceived risk, mean (SD)					
Perceived susceptibility	2.63 (0.8)	2.62 (0.8)	2.62 (0.8)	2.63 (0.9)	.78[c]
Perceived severity	3.68 (0.9)	3.68 (0.9)	3.72 (0.9)	3.63 (0.9)	.16[c]
Response efficacy, mean (SD)					
Wearing facial masks	3.69 (0.5)	3.69 (0.5)	3.76 (0.5)	3.62 (0.6)	<.001[c]
Hand hygiene	3.78 (0.5)	3.77 (0.5)	3.83 (0.4)	3.72 (0.5)	.002[c]
Keeping away from crowded places	3.67 (0.6)	3.66 (0.58)	3.76 (0.5)	3.57 (0.6)	<.001[c]
Cancelling or postponing social events	3.70 (0.6)	3.70 (0.6)	3.77 (0.5)	3.62 (0.6)	<.001[c]
Practicing precautionary behavior (“always”), n (%)					

Variables		Participants (N=990)	Weighted[a] values (N=990)	Participants who read text messages		*P* value[b,c]
				Always (n=487)	Otherwise (n=503)	
	Wearing facial masks	762 (77.1)	760 (76.8)	400 (82.4)	359 (71. 5)	<.001[b]
	Hand hygiene	716 (72.3)	714 (72.1)	383 (78.8)	330 (65.6)	<.001[b]
	Keeping away from crowded places	574 (58)	571 (57.7)	325 (66.9)	246 (48.9)	<.001[b]
	Cancelling or postponing social events	631 (63.7)	628 (63.4)	354 (72.9)	273 (54.3)	<.001[b]

[a]Data were weighted by sex, age, and regional distribution of the population in South Korea.

[b]*P* values for chi-square test.

[c]*P* values for *t* test.

[d]A currency exchange conversion rate of KRW 1=US $ 0.00091 is applicable.

Hierarchical linear regressions were computed to estimate the role of reading emergency alert text messages toward each precautionary behavior practiced against COVID-19 by sequentially adding predictors into 3 blocks within each model. To control for the effects of covariates on the dependent variables, sociodemographic factors (ie, sex, age, education level, income level, occupation status, presence of children, and underlying disease), media exposure, and obtaining information were entered into block 1 as potential confounding factors affecting each type of precautionary behavior. Psychological factors such as perceived susceptibility, perceived severity, and response efficacy of each behavior were entered into block 2 of each model. Because we were primarily interested in the effects of reading emergency alert text messages beyond these covariates, predictor variables were subsequently entered into block 3 of each model. To determine whether reading text messages resulted in any significant increment in the amount of variance explained in practicing each precautionary behavior, *F* test statistics were evaluated to determine statistically significant R^2 changes in the explained variance (%) at each step of the analysis. Standardized β coefficients were examined for each variable, and variance inflation factor estimates for each hierarchical model were computed to ensure that tolerance estimates were below 0.10 and variance inflation factor estimates were less than 10.

Additionally, the indirect effects of reading the text messages on each precautionary behavior via psychological factors (ie, perceived risk and response efficacy) were calculated using PROCESS macro model 6 with 5000 bootstrap samples for SPSS (version 25; IBM Corp) [33]. The bias-corrected 95% CI values for each mediational path were obtained.

Results

Characteristics of Survey Participants

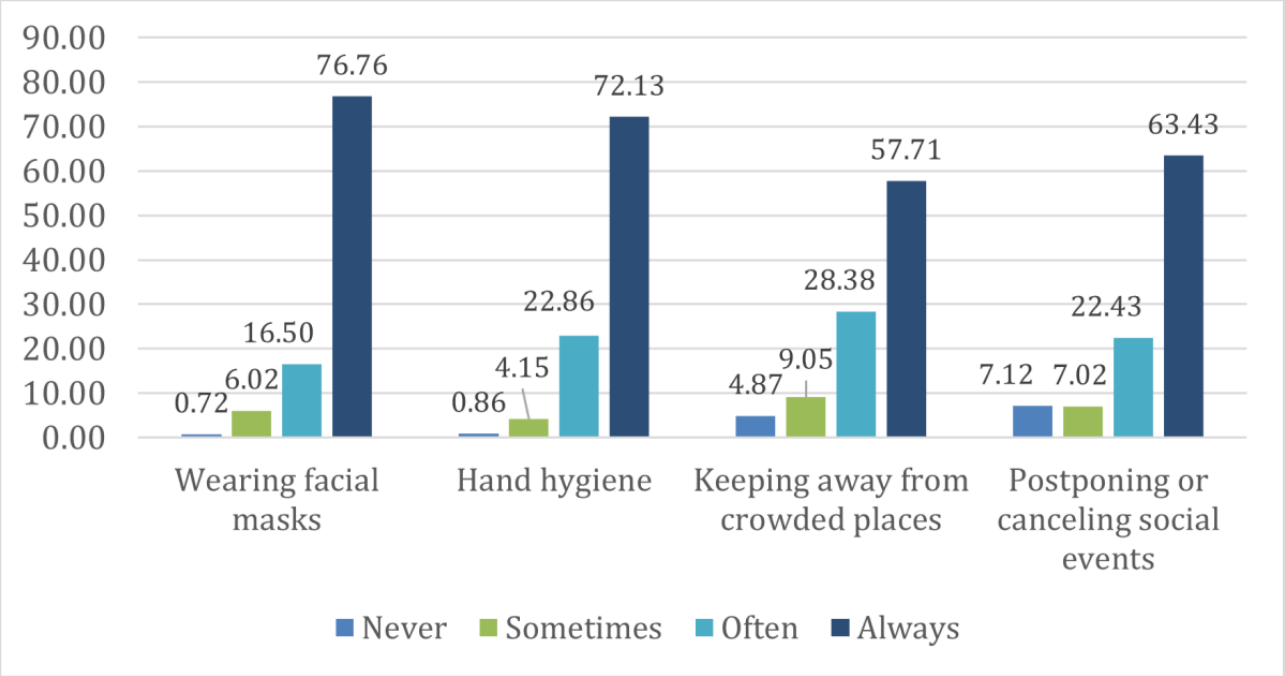
Among the 990 participants, 491 (49.6%) were men and 499 (50.4%) were women, with a mean age of 46.45 (SD 15.05) years. Most participants had received at least some college education (486/990, 49.1%), followed by those with a high school education (475/990, 48%). The most common monthly household income was approximately KRW 2-3.99 million (US $1800-3600), followed by more than KRW 6 million (US $5400) and KRW 4-5.99 million (US $3600-5400). With regard to their occupation status, 60.9% (602/990) of the participants were working, and 39.1% (387/990) were out of labor. Moreover, 9.7% (96/990) of the participants had young children at home, and 40.4% (399/990) of the participants reported having more than one underlying disease (Table 2).

Psychological Factors and Precautionary Behaviors Related to COVID-19

Participants' perceived risk of contracting COVID-19 infection was measured by a 5-point Likert scale, and the average perceived susceptibility was higher than "low" (score=2; mean 2.62, SD 0.84). Only 1.7% of the participants reported a perceived chance of infection as "very high" (score=5), and 8.7% of them reported it as "high" (score=4). The majority of participants (490/990, 49.5%) reported that their chance of infection was "neither high nor low." The average perceived severity score was higher than the perceived susceptibility score, which was close to "high" (score=4; mean 3.68, SD 0.93). However, 42.3% of the participants reported perceived severity as "high" (score=4), and 18.7% of them reported it to be "very high" (score=5). With regard to response efficacy, which was measured on a 4-point Likert scale for the four precautionary behaviors, the response efficacy of practicing hand hygiene was the highest (mean 3.77, SD 0.49), followed by that of canceling or postponing social events (mean 3.70, SD 0.57), wearing facial masks (mean 3.69, SD 0.54), and avoiding crowded places (mean 3.66, SD 0.58), as shown in Table 2.

The most frequently practiced precautionary behavior was wearing a facial mask when outside, for which 76.8% (760/990) reported they "always" practiced the behavior. For hand hygiene, such as washing hands frequently and using hand sanitizers, 72.1% (714/990) reported they "always" practiced the behavior. With regard to social distancing behaviors, postponing or canceling social events was the most practiced behavior (628/990, 63.4% reporting "always"), followed by avoiding crowded places (571/990, 57.7% reporting "always"), as shown in Figure 2.

Figure 2. Frequency of practicing precautionary behaviors in the past week (N=990).



Reading Emergency Alert Text Messages

For reading COVID-19–related emergency alert text messages, 49.2% (487/990) of the participants reported they *always* read the message and visited the linked website to obtain more information. Additionally, 40.3% (399/990) of the participants reported they read the message occasionally and visited the linked website if necessary, 8.4% (83/990) reported they rarely read the message, and 1.9% (21/990) reported they never read the message. Among those who reported they rarely or never read the emergency alert text messages (n=104), most reported not reading the messages because they were “sent too often” (73/104, 70.2%), followed by those who reported “did not want to know the information“ (12/104, 11.5%) or “did not need the information” (9/104, 8.6%).

A multivariable logistic analysis was performed to examine the factors that influence reading emergency alert text messages. Reading text messages was positively associated with female sex (odds ratio [OR] 1.68, 95% CI 1.28-2.21; $P<.001$) and older age (30-39 years: OR 2.02, 95% CI 1.25-3.28; $P<.001$; 40-49 years: OR 2.84, 95% CI 1.80-4.47; $P<.001$; 50-59 years: OR 3.19, 95% CI 2.01-5.06; $P<.001$; ≥60 years: OR 3.12, 95% CI 2.00-4.86; $P<.001$ versus 18-29 years). Men in their 20s comprised the primary sample group that ignored the emergency alert text messages (Table 3).

Table 3. Sociodemographic factors related to reading emergency alert text messages.

Variables	Participants who read text messages (Always=1, otherwise=0)		
	Odds ratio	95% CI	*P* value
Sex			
Male	Ref[a]	Ref	
Female	1.68	1.28-2.21	<.001
Age group (years)			
18-29	Ref	Ref	
30-39	2.02	1.25-3.28	<.001
40-49	2.84	1.80-4.47	<.001
50-59	3.19	2.01-5.06	<.001
≥60	3.12	2.00-4.86	<.001
Education level			
Middle school or below	Ref	Ref	
High school graduate	1.91	0.83-4.36	.13
College and above	2.10	0.91-4.88	.08
Monthly household income (million KRW)[b]			
<2	Ref	Ref	
2-3.99	0.89	0.57-1.38	.60
4-5.99	0.89	0.56-1.41	.61
≥6	0.67	0.42-1.08	.10
Occupation status			
Out of labor	Ref	Ref	
Working	0.94	0.69-1.27	.69
Presence of children			
None	Ref	Ref	
More than 1	0.94	0.59-1.50	.79
Underlying disease			
None	Ref	Ref	
More than 1	0.95	0.71-1.27	.71
Cox-Snell R^2	N/A[c]	N/A	0.06
Nagelkerke R^2	N/A	N/A	0.08

[a]Ref: reference value.

[b]A currency exchange conversion rate of KRW 1=US $ 0.00091 is applicable.

[c]N/A: not applicable.

Factors Influencing Practicing Preventive Behaviors

Hierarchical linear regression models were used to test the association of factors influencing practicing preventive behaviors, including perceived risk and response efficacy, and reading emergency alert text messages (Table 4). For wearing facial masks, sociodemographic factors, media exposure, and information seeking (step 1) accounted for 7.4% of the variance and psychological factors (step 2) explained an additional 17.9% of the variance. Adding reading text messages as predictor variable (step 3) explained an additional 0.5% variance for the behavior wearing facial masks (F_{975}=26.05, P<.001). After adjusting for potentially confounding factors, reading text messages were found to be positively associated with wearing facial masks (β=.074, P=.012). Overall, the effect of response efficacy of practicing this behavior was found to be significant and the strongest (β=.431, P<.001). The positive effect of female sex (β=.108, P<.001) and younger age (β=−.064, P=.048) were also found to be significant.

Table 4. Influencing factors associated with practicing preventive behaviors.

Variables	Wearing facial masks						Hand hygiene					
	Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
	β[a]	*P* value	β[a]	*P* value	β[a]	*P* value	β[a]	*P* value	β[a]	*P* value	β[a]	*P* value
Sex	.201	<.001	.117	<.001	.108	<.001	.181	<.001	.108	<.001	.101	.001
Age	−.086	.02	−.056	.08	−.064	.048	−.056	.12	−.052	.11	−.057	.08
Education level	.026	.43	.016	.58	.016	.59	.047	.15	.023	.44	.023	.45
Monthly household income	−.004	.90	−.002	.96	.003	.91	.004	.90	.011	.71	.014	.63
Occupation status	−.088	.008	−.053	.08	−.055	.07	−.033	.33	−.007	.81	−.008	.78
Presence of children	−.004	.90	−.005	.86	−.005	.85	−.002	.96	−.012	.68	−.012	.68
Underlying disease	−.037	.28	−.025	.41	−.024	.44	−.034	.33	−.016	.61	−.015	.64
Media exposure	.048	.18	.006	.86	−.004	.89	.133	0	.075	.02	.069	.03
Information seeking	.086	.01	.042	.19	.028	.38	.103	.003	.063	.046	.054	.095
Perceived susceptibility	—[b]	—	.003	.93	.005	.87	—	—	.017	.57	.018	.53
Perceived severity	—	—	.03	.301	.029	.33	—	—	.054	.06	.053	.07
Response efficacy	—	—	.436	<.001	.431	<.001	—	—	.431	<.001	.429	<.001
Reading text messages	—	—	—	—	.074	.01	—	—	—	—	.05	.09
R^2	0.074	—	0.253	—	0.258	—	0.074	—	0.252	—	0.254	—
ΔR^2	—	—	0.179	<.001	0.005	.01	—	—	0.178	<.001	0.002	0.09
F (*df*)	8.677 (979)	<.001	27.543 (976)	<.001	26.05 (975)	<.001	8.692 (979)	<.001	27.465 (976)	<.001	25.618 (975)	<.001

[a]Standardized β coefficients are reported.

[b]Not applicable.

For hand hygiene, sociodemographic factors, media exposure, and information seeking (step 1) accounted for 7.4% of the variance and psychological factors (step 2) explained an additional 17.8% of the variance. Reading text messages (step 3) did not account for significant additional variance in practicing hand hygiene ($\Delta R^{2=}$0.002, P=.09; F_{975}=25.618, P<.001). Response efficacy (β=.429, P<.001) was found to be significantly associated among the psychological factors. Furthermore, the coefficient of always reading text messages was not significant to practicing hand hygiene.

In terms of the social distancing behavior of keeping away from crowded places, sociodemographic factors, media exposure, and information seeking (step 1) accounted for 5.6% of the variance. Adding psychological factors (step 2) explained an additional 8.9% of the variance. Reading text messages (step 3) explained an additional 0.5% of the variance(F_{975}=13.253, P<.001). Among psychological factors, only response efficacy was found to be associated with the behavior of avoiding crowded places (β=.299, P<.001). Reading emergency alert text messages had a significantly positive effect, and participants who reported they always read text messages were more likely to practice avoiding crowded places (β=.078, P=.01).

For the social distancing behavior of cancelling or postponing social gatherings, sociodemographic factors, media exposure, and information seeking (step 1) accounted for 3.7% of the variance, and psychological factors (step 2) explained an additional 9.6% of the variance. Adding reading text messages as predictor variable (step 3) explained an additional 0.9% of the variance in wearing facial masks (F_{975}=12.425, P<.001). After adjusting for potentially confounding factors, reading text messages was found to be positively associated with canceling or postponing social gatherings (β=.103, P=.001). The effect of response efficacy of practicing this behavior was significant and the strongest (β=.306, P<.001), as shown in Table 5.

Table 5. Influencing factors associated with practicing social distancing behaviors.

Variables	Keeping away from crowded places						Canceling or postponing social events					
	Model 1		Model 2		Model 3		Model 1		Model 2		Model 3	
	β[a]	*P* value	β[a]	*P* value	β[a]	*P* value	β[a]	*P* value	β[a]	*P* value	β[a]	*P* value
Sex	.079	.02	.03	.35	.021	.52	.064	.06	.024	.46	.011	.74
Age (year)	.005	.89	.004	.91	−.004	.91	.018	.63	.023	.51	.013	.71
Education level	.029	.38	.039	.22	.038	.23	.03	.38	.043	.18	.042	.19
Monthly household income	.03	.37	.023	.47	.028	.37	.03	.37	.012	.70	.02	.54
Occupation status	−.052	.12	−.052	.11	−.053	.095	−.051	.13	−.038	.24	−.041	.21
Presence of children	.018	.57	.007	.82	.007	.82	−.005	.87	−.025	.41	−.025	.41
Underlying disease	.005	.88	.022	.51	.023	.48	.008	.82	.022	.52	.024	.48
Media exposure	.087	.02	.057	.09	.047	.17	.074	.04	.035	.30	.022	.53
Information seeking	.147	<.001	.102	.003	.088	.01	.105	.003	.06	.08	.041	.23
Perceived susceptibility	—[b]	—	.007	.82	.009	.77	—	—	.026	.40	.029	.35
Perceived severity	—	—	.026	.39	.025	.42	—	—	.055	.08	.053	.09
Response efficacy	—	—	.306	<.001	.299	<.001	—	—	.313	<.001	.306	<.001
Reading text messages	—	—	—	—	.078	.01	—	—	—	—	.103	.001
R^2	.056	—	.145	—	.15	—	.037	—	.133	—	.142	—
ΔR^2	—	—	.089	<.001	.005	.01	—	—	.096	<.001	.009	.001
F (*df*)	6.462 (979)	<.001	13.779 (976)	<.001	13.253 (975)	<.001	4.152 (979)	<.001	12.47 (976)	<.001	12.425 (975)	<.001

[a]Standardized β coefficients are reported.

[b]Not applicable.

Mediation Effect of Psychological Factors Between Reading Text Messages and Practicing Preventive Behaviors

We examined the direct and indirect effects of reading text messages that were mediated via psychological factors such as perceived risk and response efficacy of practicing behaviors. When it comes to the recommended behavior, the direct and indirect effects of reading text messages on wearing masks was significant. For hand hygiene, only the indirect effect, mediated via response efficacy of the behaviors, was significant. Concerning social distancing, the direct and indirect effects of reading text messages mediated by the response efficacy of the behavior were also significant (Table 6). Therefore, this result confirms the significant effect of reading text messages on practicing precautionary behavior.

Table 6. Direct and indirect effects of reading emergency alert text messages on practicing preventive behaviors based on perceptions (eg, perceived susceptibility, severity, and response efficacy). Unstandardized point estimates represent the indirect effect of the independent variable on the dependent variable through the mediator.

Dependent variable	Preventive behaviors		Social distancing behaviors	
	Wearing facial masks	Hand hygiene	Keeping away from crowded places	Canceling or postponing social events
	Estimate (95% CI[a])	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)
Total effect	0.1257 (0.0475 to 0.2039)	0.0795 (0.0031 to 0.1558)	0.1772 (0.0682 to 0.2861)	0.2190 (0.1020 to 0.3361)
Direct effect	0.0887 (0.0178 to 0.1595)	0.0602 (−0.0088 to 0.1292)	0.1276 (0.0228 to 0.2323)	0.1804 (0.0684 to 0.2924)
Indirect effect (via perceived susceptibility)	−0.0001 (−0.0039 to 0.0035)	−0.0006 (−0.0047 to 0.0027)	−0.0004 (−0.0061 to 0.0043)	−0.0014 (−0.0084 to 0.0037)
Indirect effect (via perceived severity	0.0005 (−0.0029 to 0.0047)	0.0009 (−0.0036 to 0.0063)	0.0006 (−0.0039 to 0.0063)	0.0012 (−0.0046 to 0.0095)
Indirect effect (via response efficacy)	0.0366 (0.0041 to 0.0705)	0.019 (0.0134 to 0.0508)	0.0494 (0.0161 to 0.0847)	0.0389 (0.0043 to 0.0769)

[a]Bias-corrected CI (these 95% CIs do not cross zero; thus, mediation is assumed).

Discussion

Principal Findings

In this study, the effects of reading emergency alert text messages on practicing preventive behaviors during the COVID-19 outbreak in South Korea were examined. Reading text messages was found to be related to all precautionary behaviors tested, including wearing facial masks, practicing hand hygiene, avoiding crowded places, and avoiding social gatherings. Reading text messages directly or indirectly affected practicing precautionary behaviors, as the mediation effect of response efficacy between reading text messages and practicing those behaviors was significant. Sociodemographic factors (eg, male participants in their 20s) were related to a lower likelihood of reading emergency alert text messages.

Several findings provide valuable insights to the public health management authorities to prevent and protect the population's health during an emerging infectious disease outbreak. First, our study results indicate that health risk communication via emergency alert text messages is efficient and effective for engaging the public in practicing preventive behaviors during public health emergencies. Participants who always read and visited the linked website practiced precautionary behaviors 1.48-1.80 times more frequently than other participants. This effect was significant, even when media use was included in the logistic model as a covariate. Early release of official guidelines and timely provision of information led by governments to guide the public on responding to emergencies is essential during public health emergencies [34,35]. Therefore, mobile phones, especially text messages, can be an effective means of communication to elicit the rapid response required by public health authorities.

The phenomenon of “infodemics,” defined as the rapid spread and amplification of vast amounts of valid and invalid information on the internet or through other media, is a tremendous and ongoing challenge on the COVID-19 pandemic [36,37]. Since the beginning of the COVID-19 pandemic, both the production and consumption of information have increased rapidly and significantly [9,12]. According to a study conducted in Korea, more than two-thirds of the participants reported exposure to COVID-19 misinformation between January and April 2020 [38]. Communication via emergency alert text messages can be an effective strategy for public health authorities to provide accurate and reliable information, confront misinformation or disinformation, and reduce the negative impact of such infodemics. Additionally, digital inequality, also known as the digital divide, is a significant concern among national and international scholars and policy makers, as internet access such as material, skills, and usage is not evenly distributed among the general population [39,40]. People can receive text messages on their mobile phones without using a smartphone or connecting to the internet. Therefore, text messages can be a communication tool to alleviate digital inequalities.

Second, some people rarely or never read text messages. The present study identified a subpopulation that does not pay attention to the text messages sent by the public health authorities. People often ignore the emergency warning and alert messages as initial responses are often marked by skepticism, disbelief, and denial [41-43]. Explanations for the lack of response include repeated, long-term exposure to a warning that may contribute to message fatigue and a loss of ability to capture and maintain the attention needed to elicit a response at later times [25,44-46]. Kuligowski and Dootson [35] proposed a 6-step process to process information and respond. These steps include (1) receiving the information, (2) paying attention to the message, (3) aiming to understand the information provided, (4) believing the threat presented by the message, (5) personalizing the presented risk to themselves or others, and finally, (6) deciding to take protective action and to respond based on the information received in the message

[19,35,47]. Their study indicates that a high level of attention to the provided information is essential for the recipient to move to the next step. Hence, convincing the public that reading emergency alert text messages can help guide them to better respond to public health emergencies is important.

Considering the COVID-19 emergency has lasted about a year and seems likely to continue for some time, more people will likely not pay attention to text messages in the future. Therefore, further research is needed to identify the reasons for not paying attention to text messages and investigate how to maintain interest. The content, length, frequency, and timing of the alert should be reviewed to maximize the effectiveness of emergency alert text messages. Moreover, people with limited literacy might resist to read or be attentive toward these text messages. Using a symbol and number-coded system that is easy to remember and interpret by the public, even among those with limited literacy, can be considered to address this challenge [48]. As communication can help reduce health inequalities derived from health communication [49], efforts to reduce vulnerability should also be implemented.

This study also identified a subpopulation that read alert text messages less frequently than others. Men in their 20s were the least likely to read the text messages they received. Infrequent reading of text messages by participants in their 20s might be because they are more proficient in using mobile technology and prefer obtaining information via online digital resources. However, their digital health literacy—the ability to evaluate health resources and apply gathered information to health-related decisions—was relatively low [38,50]. Therefore, young people have less opportunity to obtain official information from public health authorities and are more likely to be exposed to misinformation distributed online. This disparity can make them more vulnerable to the COVID-19 infodemic. Additionally, current practice in emergency alert text messages sends the same message, regardless of the recipient's demographic characteristics. To enhance the efficacy of text messages, tailored messages for each population and subgroups can be considered.

Finally, reading text messages in this study was mediated through heightened response efficacy of practicing behaviors rather than perceived risk. Moreover, response efficacy is one factor that has the strongest positive effect on practicing preventive behaviors [13,51,52]. Public health authorities and policymakers can consider making efforts to improve messages sent via texting to strengthen response efficacy on practicing preventive behaviors. For instance, investigating the unmet needs of health risk information and providing for them can be helpful. In other words, receiver-centered messages should be provided. We propose that future studies analyze the messages sent by texting to design more effective messages.

Implications

A set of implications for interventions, communication strategies, and future research can be drawn from the findings of this study. Governments, health agencies, and researchers should take advantage of mobile text messages by producing and sharing evidence-based and correct information, as well as recommending precautionary behaviors to the public. Public health authorities should pay careful attention to what messages are provided to the public. Based on the findings of this study, we believe that persuasive messages that target to improve response efficacy on recommended behaviors would be most useful.

Second, for effective text message communication, the public needs a moderate or higher trust level in the government and health authorities [48]. Successful risk communication and efficacy of policy recommendations depend not only on how quickly or reliably a message can be disseminated but also on the individuals' beliefs and subsequent response [6,15,16,53,54]. The first step in this process is the firm belief of whether the recommended action will mitigate the threat or manage the fearful situation [54].

Third, health authorities' efforts to provide transparent and credible information should be encouraged and sustained. South Korea has prior experience in managing outbreaks of infectious diseases such as the MERS outbreak in 2015, which resulted in significant damage to the Korean population, widespread distrust, and societal levels of high stress. Nevertheless, the MERS pandemic has provided many important lessons, especially the importance of information disclosure to the public [13]. In case of the MERS outbreak, the Korean health authorities' decision to disclose specific information, such as hospitals that were exposed to MERS-CoV, contributed to further prevention of the spread of infection during the outbreak [55].

Limitations

The present study has several limitations. First, variables in this study were assessed using a single-item questionnaire. Single items are less time-consuming, minimize participant burden, and prove beneficial to surveys that need to be done quickly to provide timely information. Single-item measures are argued to have comparable or equal predictive validity compared to multiple-item measures in some study areas [56]. However, this approach can be problematic because of the unknown biases in meaning and interpretation, and the internal consistency, as well as reliability statistics, cannot be tested. Therefore, additional studies using multi-item questionnaires should be performed.

Second, the participants did not all receive the same text messages. As shown in Table 1, the central governments' messages were the same for all of the population, and the contents were mostly about public health policy updates, persuading practice of precautionary behaviors. The messages sent from local governments usually covered more residence-specific risk information, such as reports of confirmed cases in the area, contact tracing of confirmed cases, and closure of certain places. This study was not designed to test the effect of specific messages, like previous experimental studies [57-59]. There is still a chance that differences in how people are primed might influence their preventive behaviors. Nevertheless, this study has the advantage of being conducted on many people in real-life situations. This study's results should be interpreted focusing on the effect of using emergency alert text messages as one of the information sources, rather than focusing on the effect of the content of the messages received.

Third, we did not explore self-efficacy in this study. Self-efficacy refers to the individual's level of confidence in preventing the risk and is an essential coping appraisal component [54,60]. Thus, further study to examine the influence of self-efficacy is suggested. Lastly, the participants ' characteristics, such as health literacy, government trust, or perceived credibility on the text messages, were not explored in this study. These variables may contribute to both whether and how messages were read and whether precautionary behaviors were practiced. The effects of these variables should also be examined in further study.

Conclusions

New information technologies, communication devices, apps, and social media for health risk communication are continually emerging. However, sometimes *oldies are goodies*, which this study implicates. The present study provides evidence that emergency alert text messages sent to an individual's mobile phone are efficient and an effective communication strategy for the sustainability of preventive behaviors among the public. Government and public health authorities should use text messages to provide the population with accurate and reliable information. This approach can reduce the negative impact of an infodemic by confronting misinformation or disinformation. At the same time, efforts to ensure people keep reading text messages should be implemented. This study provides critical and timely insights into how governments and public health authorities build appropriate health risk communications that do not overlook and lower their priorities for those in urgent need.

Acknowledgments

This work was supported by the National Research Foundation of Korea (BK21 Center for Integrative Response to Health Disasters, Graduate School of Public Health, Seoul National University); National Research Foundation of Korea Grant funded by the Korean Government (No. 21B20151213037); and Rainbow Communication and Consulting.

Authors' Contributions

LMJ and YMS conceptualized the study. LMJ was responsible for the methodology and conducted a formal analysis. YMS was responsible for data acquisition. LMJ wrote the initial draft of the manuscript, and YMS assisted with writing, review, and editing of the manuscript. Both authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Eiselt H, Marianov V. Mobile phone tower location for survival after natural disasters. European Journal of Operational Research 2012 Feb;216(3):563-572. [doi: 10.1016/j.ejor.2011.08.021]
2. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
3. Meier P. The unprecedented role of SMS in disaster response: learning from Haiti. SAIS Review of International Affairs 2010;30(2):91-103. [doi: 10.1353/sais.2010.0009]
4. Chu HS, Tak YR, Lee H. Exploring psychosocial factors that influence smartphone dependency among Korean adolescents. PLoS One 2020;15(5):e0232968 [FREE Full text] [doi: 10.1371/journal.pone.0232968] [Medline: 32401789]
5. Glik D, Eisenman D, Johnson K, Prelip M, Arevian A, Martel A, et al. WEA messages: Impact on physiological, emotional, cognitive and behavioral responses. WEA messages: impact on physiological, emotional, cognitive and behavioral responses. Los Angeles, CA: Department of Homeland Security Science and Technology, UCLA Fielding School of Public Health, Center for Public Health and Disasters; 2016. URL: https://www.dhs.gov/sites/default/files/publications/UCLA%20-%20WEA%20Final%20Report%2008-11-2016-508.pdf [accessed 2021-01-10]
6. Lee D, Chung JY, Kim H. Text me when it becomes dangerous: Exploring the determinants of college students' adoption of mobile-based text alerts short message service. Computers in Human Behavior 2013 May;29(3):563-569. [doi: 10.1016/j.chb.2012.11.014]
7. Zhang N, Huang H, Su B, Zhao J, Zhang B. Information dissemination analysis of different media towards the application for disaster pre-warning. PLoS One 2014;9(5):e98649 [FREE Full text] [doi: 10.1371/journal.pone.0098649] [Medline: 24878739]
8. Ur Rehman A, Lyche T, Awuah-Offei K, Nadendla V. Effect of text message alerts on miners evacuation decisions. Safety Science 2020 Oct;130:104875. [doi: 10.1016/j.ssci.2020.104875]
9. Wong DJ, Jones E, Rubin GJ. Mobile text alerts are an effective way of communicating emergency information to adolescents: Results from focus groups with 12- to 18-year-olds. J Contingencies Crisis Man 2017 Jul 18;26(1):183-192. [doi: 10.1111/1468-5973.12185]

10. Intensive social distancing, the participation of employees is required! (Webpage in Korean).: Korea Centers for Disease Control and Prevention; 2020 Mar 23. URL: https://m.post.naver.com/viewer/postView.nhn?volumeNo=27795996&memberNo=20182790 [accessed 2021-02-16]
11. Corona 19 Word of the Day (3.24.) (Webpage in Korean).: Korea Centers for Disease Control and Prevention; 2020 Mar 24. URL: https://m.post.naver.com/viewer/postView.nhn?volumeNo=27816495&memberNo=20182790 [accessed 2021-02-16]
12. Rubin GJ, Chowdhury AK, Amlôt R. How to communicate with the public about chemical, biological, radiological, or nuclear terrorism: a systematic review of the literature. Biosecur Bioterror 2012 Dec;10(4):383-395. [doi: 10.1089/bsp.2012.0043] [Medline: 23216210]
13. Lee M, You M. Psychological and behavioral responses in South Korea during the early stages of coronavirus disease 2019 (COVID-19). IJERPH 2020 Apr 25;17(9):2977. [doi: 10.3390/ijerph17092977]
14. Drake JM, Chew SK, Ma S. Societal learning in epidemics: intervention effectiveness during the 2003 SARS outbreak in Singapore. PLoS One 2006 Dec 20;1:e20 [FREE Full text] [doi: 10.1371/journal.pone.0000020] [Medline: 17183647]
15. Dash N, Gladwin H. Evacuation decision making and behavioral responses: individual and household. Nat Hazards Rev 2007 Aug;8(3):69-77. [doi: 10.1061/(asce)1527-6988(2007)8:3(69)]
16. Fakhruddin S, Chivakidakarn Y. A case study for early warning and disaster management in Thailand. International Journal of Disaster Risk Reduction 2014 Sep;9:159-180. [doi: 10.1016/j.ijdrr.2014.04.008]
17. Ajzen I. The theory of planned behavior. Organizational behavior and human decision processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
18. Lindell MK, Perry RW. In: Gudykunst WB, Ting-Toomey S, editors. Communicating environmental risk in multiethnic communities. Sage Publications: Thousand Oaks, CA; 2004.
19. Mileti DS, Peek L. The social psychology of public response to warnings of a nuclear power plant accident. J Hazard Mater 2000 Jul 28;75(2-3):181-194. [doi: 10.1016/s0304-3894(00)00179-5] [Medline: 10838242]
20. Rogers RW. A protection motivation theory of fear appeals and attitude change1. J Psychol 1975 Sep;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
21. Mulilis J, Duval TS, Rogers R. The effect of a swarm of local tornados on tornado preparedness: a quasi-comparable cohort investigation. J Appl Social Pyschol 2003 Aug;33(8):1716-1725. [doi: 10.1111/j.1559-1816.2003.tb01971.x]
22. Bean H, Liu BF, Madden S, Sutton J, Wood MM, Mileti DS. Disaster warnings in your pocket: how audiences interpret mobile alerts for an unfamiliar hazard. J Contingencies & Crisis Man 2016 Mar 09;24(3):136-147. [doi: 10.1111/1468-5973.12108]
23. Casteel M, Downing J. Assessing risk following a wireless emergency alert: Are 90 characters enough? Journal of Homeland Security Emergency Management 2016;13(1):A. [doi: 10.1515/jhsem-2015-0024]
24. Sorensen JH. Hazard warning systems: review of 20 years of progress. Nat Hazards Rev 2000 May;1(2):119-125. [doi: 10.1061/(asce)1527-6988(2000)1:2(119)]
25. Kim G, Martel A, Eisenman D, Prelip M, Arevian A, Johnson KL, et al. Wireless emergency alert messages: influences on protective action behaviour. J Contingencies and Crisis Management 2019 Sep 10;27(4):374-386. [doi: 10.1111/1468-5973.12278]
26. Liu BF, Fraustino JD, Jin Y. Social media use during disasters. Communication Research 2015 Jan 13;43(5):626-646. [doi: 10.1177/0093650214565917]
27. Nulty DD. The adequacy of response rates to online and paper surveys: what can be done? Assessment & Evaluation in Higher Education 2008 Jun;33(3):301-314. [doi: 10.1080/02602930701293231]
28. Fincham JE. Response rates and responsiveness for surveys, standards, and the Journal. Am J Pharm Educ 2008 Apr 15;72(2):43 [FREE Full text] [doi: 10.5688/aj720243] [Medline: 18483608]
29. Eysenbach G, Wyatt J. Using the internet for surveys and health research. J Med Internet Res 2002;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
30. Slovic P. Perception of risk. Science 1987 Apr 17;236(4799):280-285. [doi: 10.1126/science.3563507] [Medline: 3563507]
31. Rimal RN. Perceived risk and efficacy beliefs as motivators of change: use of the Risk Perception Attitude (RPA) framework to understand health behaviors. Human Communication Research 2003 Jul 01;29(3):370-399. [doi: 10.1093/hcr/29.3.370]
32. KOSIS National population cencus. Korean Statistical Information Service. 2019. URL: https://kosis.kr/index/index.do [accessed 2021-02-15]
33. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York, NY: Guilford Publications; 2017.
34. Hou Z, Du F, Zhou X, Jiang H, Martin S, Larson H, et al. Cross-country comparison of public awareness, rumors, and behavioral responses to the COVID-19 epidemic: infodemiology study. J Med Internet Res 2020 Aug 03;22(8):e21143 [FREE Full text] [doi: 10.2196/21143] [Medline: 32701460]
35. Kuligowski E, Dootson P. Emergency Notification: Warnings and Alerts. In: Manzello S, editor. Encyclopedia of Wildfires Wildland-Urban Interface Fires. Cham: Springer; 2018.
36. Ashrafi-Rizi H, Kazempour Z. Information typology in coronavirus (COVID-19) crisis; a commentary. Arch Acad Emerg Med 2020;8(1):e19 [FREE Full text] [Medline: 32185370]

37. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
38. Lee JJ, Kang K, Wang MP, Zhao SZ, Wong JYH, O'Connor S, et al. Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and preventive behaviors: cross-sectional online study. J Med Internet Res 2020 Nov 13;22(11):e22205 [FREE Full text] [doi: 10.2196/22205] [Medline: 33048825]
39. van Deursen AJ. Digital inequality during a pandemic: quantitative study of differences in COVID-19-related internet uses and outcomes among the general population. J Med Internet Res 2020 Aug 20;22(8):e20073 [FREE Full text] [doi: 10.2196/20073] [Medline: 32750005]
40. Scheerder A, van Deursen A, van Dijk J. Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telematics and Informatics 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
41. Drabek TE. Social processes in disaster: family evacuation. Social Problems 1969 Jan;16(3):336-349. [doi: 10.2307/799667]
42. Ferrell WR, Krzysztofowicz R. A model of human response to flood warnings for system evaluation. Water Resour Res 2010 Jul 09;19(6):1467-1475. [doi: 10.1029/wr019i006p01467]
43. Vermeulen K. Understanding your audience: How psychologists can help emergency managers improve disaster warning compliance. Journal of Homeland Security Emergency Management 2014;11(3):309-315. [doi: 10.1515/jhsem-2014-0055]
44. Woody C, Ellison R. Maximizing trust in the wireless emergency alerts (WEA) service. Carnegie Mellon University Software Engineering Institute 2014. [doi: 10.21236/ada609838]
45. Thorley P, Hellier E, Edworthy J. Habituation effects in visual warnings. Contemporary ergonomics 2001:223-230. [doi: 10.1201/b12798-36]
46. Wogalter MS, Godfrey SS, Fontenelle GA, Desaulniers DR, Rothstein PR, Laughery KR. Effectiveness of warnings. Hum Factors 2016 Nov 23;29(5):599-612. [doi: 10.1177/001872088702900509]
47. Perry R, Lindell M. Preparedness for emergency response: guidelines for the emergency planning process. Disasters 2003 Dec;27(4):336-350. [doi: 10.1111/j.0361-3666.2003.00237.x] [Medline: 14725091]
48. Sakib M, Butt Z, Morita P, Oremus M, Fong G, Hall P. Considerations for an individual-level population notification system for pandemic response: a review and prototype. JMIR Preprints. Preprint posted online on May 06, 2020 [FREE Full text] [doi: 10.2196/preprints.19930]
49. Lee M, Ju Y, You M. The effects of social determinants on public health emergency preparedness mediated by health communication: the 2015 MERS outbreak in South Korea. Health Commun 2020 Oct;35(11):1396-1406. [doi: 10.1080/10410236.2019.1636342] [Medline: 31262197]
50. Heuberger RA, Ivanitskaya L. Preferred sources of nutrition information: contrasts between younger and older adults. Journal of Intergenerational Relationships 2011 Apr;9(2):176-190. [doi: 10.1080/15350770.2011.567919]
51. Farooq A, Laato S, Islam AKMN. Impact of online information on self-isolation intention during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2020 May 06;22(5):e19128 [FREE Full text] [doi: 10.2196/19128] [Medline: 32330115]
52. Niu Z, Wang T, Hu P, Mei J, Tang Z. Chinese public's engagement in preventive and intervening health behaviors during the early breakout of COVID-19: cross-sectional study. JMIR Preprints. Preprint posted online on Jul 27, 2020 [FREE Full text] [doi: 10.2196/preprints.19995]
53. Guidry JP, Carlyle KE, LaRose JG, Perrin P, Messner M, Ryan M. Using the health belief model to analyze Instagram posts about Zika for public health communications. Emerg Infect Dis 2019 Jan;25(1):179-180 [FREE Full text] [doi: 10.3201/eid2501.180824] [Medline: 30561302]
54. Al-Hasan A, Khuntia J, Yim D. Threat, coping, and social distance adherence during COVID-19: cross-continental comparison using an online cross-sectional survey. J Med Internet Res 2020 Nov 18;22(11):e23019 [FREE Full text] [doi: 10.2196/23019] [Medline: 33119538]
55. Noh J, Yoo K, Kwon Y, Hong J, Lee Y, Park K. Effect of information disclosure policy on control of infectious disease: MERS-COV outbreak in South Korea. Int J Environ Res Public Health 2020 Jan 01;17(1):305 [FREE Full text] [doi: 10.3390/ijerph17010305] [Medline: 31906369]
56. Hoeppner BB, Kelly JF, Urbanoski KA, Slaymaker V. Comparative utility of a single-item versus multiple-item measure of self-efficacy in predicting relapse among young adults. J Subst Abuse Treat 2011 Oct;41(3):305-312 [FREE Full text] [doi: 10.1016/j.jsat.2011.04.005] [Medline: 21700411]
57. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014 Oct 17;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442] [Medline: 25326646]
58. Buis LR, Artinian NT, Schwiebert L, Yarandi H, Levy PD. Text messaging to improve hypertension medication adherence in African Americans: BPMED intervention development and study protocol. JMIR Res Protoc 2015 Jan 02;4(1):e1 [FREE Full text] [doi: 10.2196/resprot.4040] [Medline: 25565680]
59. Guo Y, Xu Z, Qiao J, Hong YA, Zhang H, Zeng C, et al. Development and feasibility testing of an mHealth (text message and WeChat) intervention to improve the medication adherence and quality of life of people living with HIV in China: pilot randomized controlled trial. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e10274 [FREE Full text] [doi: 10.2196/10274] [Medline: 30181109]

60. Bandura A, Freeman W, Lightsey R. Self-Efficacy: The Exercise of Control. New York, NY: W. H. Freeman and Company; 1997.

Abbreviations

KCDC: Korea Centers for Disease Control and Prevention
MERS: Middle East Respiratory Syndrome
OR: odds ratio

Edited by G Eysenbach, C Basch; submitted 07.09.20; peer-reviewed by A Farooq, M Fahimi, J Jiang; comments to author 18.11.20; revised version received 23.12.20; accepted 04.02.21; published 25.02.21.

Please cite as:
Lee M, You M
Effects of COVID-19 Emergency Alert Text Messages on Practicing Preventive Behaviors: Cross-sectional Web-Based Survey in South Korea
J Med Internet Res 2021;23(2):e24165
URL: https://www.jmir.org/2021/2/e24165
doi:10.2196/24165
PMID:33544691

Original Paper

Infection Control Behavior at Home During the COVID-19 Pandemic: Observational Study of a Web-Based Behavioral Intervention (Germ Defence)

Ben Ainsworth[1,2], PhD; Sascha Miller[3], MSc; James Denison-Day[3], PhD; Beth Stuart[4], PhD; Julia Groot[1], MSc; Cathy Rice[3], BSc; Jennifer Bostock[5], MA; Xiao-Yang Hu[4], PhD; Katherine Morton[3], PhD; Lauren Towler[3], MSc; Michael Moore[4], FRCGP; Merlin Willcox[4], DPhil, MRCGP; Tim Chadborn[6], PhD; Natalie Gold[6,7], DPhil; Richard Amlôt[8], PhD; Paul Little[4], MD; Lucy Yardley[3,9], PhD

[1]Department of Psychology, University of Bath, Bath, United Kingdom
[2]National Institute for Health Research Biomedical Research Centre, Faculty of Medicine, University of Southampton, Southampton, United Kingdom
[3]School of Psychology, University of Southampton, Southampton, United Kingdom
[4]Primary Care Population Sciences and Medical Education, University of Southampton, Southampton, United Kingdom
[5]Policy Research Unit, London School of Hygiene & Tropical Medicine, London, United Kingdom
[6]Public Health England Behavioural Insights, Public Health England, London, United Kingdom
[7]Centre for the Philosophy of Natural and Social Sciences, London School of Economics, London, United Kingdom
[8]Behavioural Science Team, Emergency Response Department Science and Technology, Public Health England, London, United Kingdom
[9]School of Psychological Science, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Ben Ainsworth, PhD
Department of Psychology
University of Bath
Claverton Down
Bath, BA27AY
United Kingdom
Phone: 44 01225388388
Email: ba548@bath.ac.uk

Abstract

Background: To control the COVID-19 pandemic, people should adopt protective behaviors at home (self-isolation, social distancing, putting shopping and packages aside, wearing face coverings, cleaning and disinfecting, and handwashing). There is currently limited support to help individuals conduct these behaviors.

Objective: This study aims to report current household infection control behaviors in the United Kingdom and examine how they might be improved.

Methods: This was a pragmatic cross-sectional observational study of anonymous participant data from Germ Defence between May 6-24, 2020. Germ Defence is an open-access fully automated website providing behavioral advice for infection control within households. A total of 28,285 users sought advice from four website pathways based on household status (advice to protect themselves generally, to protect others if the user was showing symptoms, to protect themselves if household members were showing symptoms, and to protect a household member who is at high risk). Users reported current infection control behaviors within the home and intentions to change these behaviors.

Results: Current behaviors varied across all infection control measures but were between *sometimes* (face covering: mean 1.61, SD 1.19; social distancing: mean 2.40, SD 1.22; isolating: mean 2.78, SD 1.29; putting packages and shopping aside: mean 2.75, SD 1.55) and *quite often* (cleaning and disinfecting: mean 3.17, SD 1.18), except for handwashing (*very often*: mean 4.00, SD 1.03). Behaviors were similar regardless of the website pathway used. After using Germ Defence, users recorded intentions to improve infection control behavior across all website pathways and for all behaviors (overall average infection control score mean difference 0.30, 95% CI 0.29-0.31).

Conclusions: Self-reported infection control behaviors other than handwashing are lower than is optimal for infection prevention, although handwashing is much higher. Advice using behavior change techniques in Germ Defence led to intentions to improve

these behaviors. Promoting Germ Defence within national and local public health and primary care guidance could reduce COVID-19 transmission.

(J Med Internet Res 2021;23(2):e22197) doi:10.2196/22197

KEYWORDS

COVID-19; novel coronavirus; behavior change; digital medicine; infection control; infectious disease; protection; digital health

Introduction

The impacts of COVID-19 must primarily be tackled through changes in behavior undertaken by individuals and societies until a vaccine becomes available. In many countries (including the United Kingdom), people with COVID-19 infection are instructed to remain at home, together with cohabiting family or other household members, to prevent transmission between households. This increases the risk of within-household virus transmission. For example, in several environments where interhousehold movement is well controlled (eg, Taiwan, Ningbo, and Shenzen [1-3]), the virus continues to proliferate within close contacts.

To interrupt these transmission pathways, individuals must adopt *personal protective behaviors* [4]. Such targeted behaviors include handwashing, disinfection of surfaces, thorough cleaning and waste disposal, social distancing within the home (where possible), and wearing situationally appropriate personal protective equipment. A recent cohort study in Beijing, China demonstrated that performing these behaviors could dramatically reduce the likelihood of household transmission, but the highest risk of transmission was prior to symptom onset (typically before such behaviors are performed) [5]. Therefore, protective behaviors should be implemented before any household members develop symptoms. There is substantial individual variation in these behaviors, which are complex, environmentally and culturally dependent, and influenced by individual attitudes and beliefs [6]. Changing such complex behaviors effectively and rapidly within the context of COVID-19 requires an approach based on behavior change theory, evidence, and extensive participatory input [7].

Specific guidance for the public on protective behaviors has been developed in many countries and is widely recommended by politicians, the media, and public health and primary care networks [8]. However, few behavioral interventions have been used to support the public in these behaviors within their homes. A systematic review by our group has found evidence of only one digital intervention to date (Germ Defence [9,10]) that demonstrably improved health outcomes in respiratory tract infections within households. Germ Defence is a mobile-friendly website that provides targeted, tailored advice about how and why users should use infection control behaviors, aiming to supplement public health guidance with evidence- and theory-based behavior change techniques [11], optimized using extensive user feedback. In a large randomized controlled trial of 20,066 people (the PRIMIT [Primary Care Randomised Trial of an Internet Intervention to Modify Influenza-Like Illness and Respiratory Infection Transmission] trial) during the previous H1N1 (swine flu) pandemic [12], those randomized to use Germ Defence had reduced frequency and severity of respiratory tract infections, and reduced transmission to household members. Germ Defence is a freely available resource, and the intellectual property is held by the University of Southampton.

Germ Defence was rapidly adapted for the COVID-19 pandemic by a team of medical, public health and behavior change experts, and public contributors. It was then disseminated through multiple pathways (primarily but not exclusively in the United Kingdom), including public health and primary care networks (eg, by texting the website link to patients via general practitioner practices), national and local press, television coverage, and social media.

In this study, we aim to:

1. Examine current infection control behaviors in UK households
2. Compare current infection control behaviors with intentions to change behavior after using Germ Defence to control infection transmission

Methods

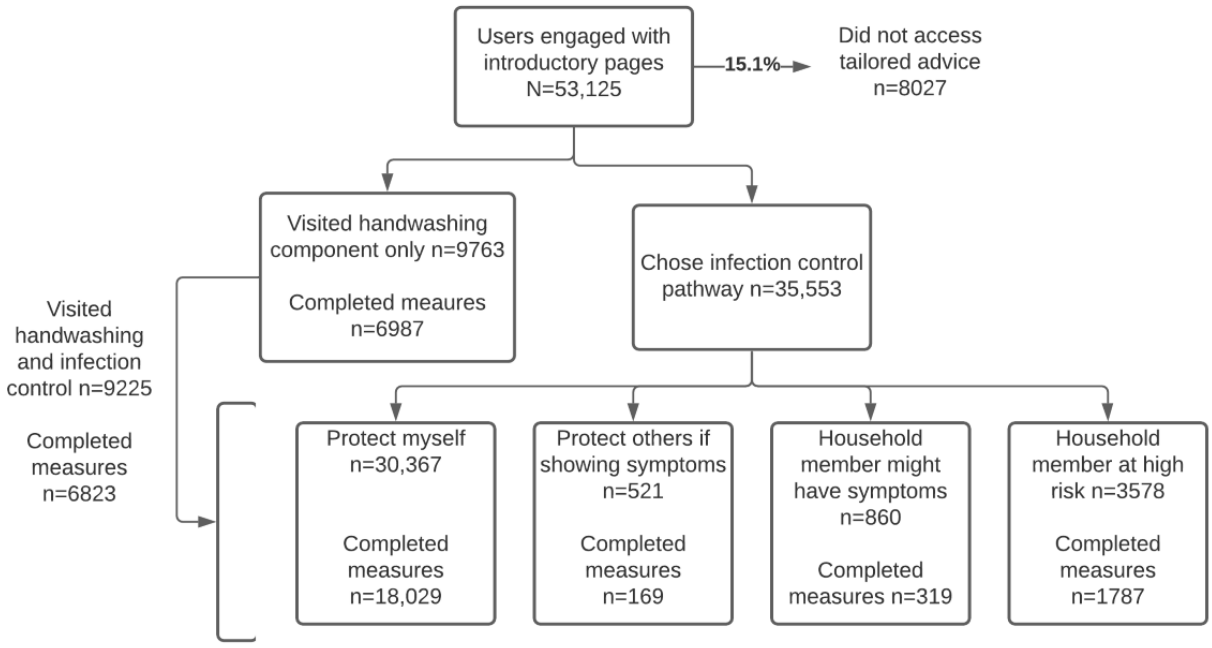
Design

This was a cross-sectional observational study of anonymous participant data from an active behavioral intervention. Consent was assumed from website use and acknowledged in the website privacy policy. All data was collected in line with General Data Protection Regulation EU Law. The study received ethical approval from the University of Bath (PREC reference 20-088). All time stamped data files used in analysis (and analysis scripts) are available at [13].

Participants and Data

The data analyzed were collected from users of the Germ Defence website between May 6 and May 24, 2020. Usage was driven by media coverage, and users were encouraged to share the intervention on social media and by email. During this period, 70,566 website hits were recorded, with 53,125 users completing the introductory content (first 3 pages) and 28,285 people completing the core module, which included measures of current and intended behavior. Website use and engagement data was collected using Google Analytics embedded in the site (see Figure 1 for full CONSORT [Consolidated Standards of Reporting Trials] use diagram).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram of Germ Defence website use and group categorization.



Data collection was kept to a minimum to reduce dropout. Behavioral measures were recorded through self-report questions within the website for current and intended behavior (see Table 1).

Table 1. Online self-report measures recorded during Germ Defence intervention

Behavior	Self-report item[a]
Reducing illness infection control	
Social distancing	When you were/are with them, how often were you/do you plan to be more than 2 meters/6 feet away from the people you live with?
Cleaning/disinfecting	How often did you/do you plan to clean things that might have viruses on them?
Putting shopping/packages aside	How often did you/do you plan to put something aside for at least 1 day that might have viruses on it?
Self-isolating	How often did you/do you plan to spend time in a room on your own?
Wearing face coverings	How often did you/do you plan to wear a face covering and glasses (and safely remove and clean them) when you are in the same room as other people?
Handwashing behavior	
Before snacking	How often did you/do you plan to wash your hands before you ate/eat with your fingers (eg, snack, fruit, or sweets)?
After coming home	How often did you/do you plan to wash your hands when you came/come into a house (eg, after work, shopping, travelling)?
After coughing	How often did you/do you plan to wash your hands after blowing your nose or sneezing/coughing on your hands?
After coming into contact with possible carrier	How often did you/do you plan to wash your hands after you had been/being close to someone who may have a virus (within 6 feet)?
After touching something	How often did you/do you plan to wash your hands after touching something that lots of other people have touched (eg, doors, money, or handrails)?
Website helpfulness (recorded on a scale of 1-10)	
Helpfulness score	How strongly do you agree or disagree that Germ Defence was helpful to you?

[a]Measures were all scored on a Likert scale with answers of 1 (almost never), 2 (sometimes), 3 (quite often), 4 (very often), and 5 (almost always). Users could also answer *not applicable* (eg, if they lived alone and therefore did not need to socially isolate within their household).

Intervention

Germ Defence content was developed using theoretical modeling and qualitative research [14] in line with the person-based approach [15], drawing principally on the theory of planned behavior [16], Leventhal's common sense model of illness [17], and protection motivation theory [18]. The intervention content, design, and structure were optimized iteratively using in-depth qualitative think-aloud interviews with public contributors (authors JB and CR) and members of the public to ensure the intervention was accessible, credible, and motivating for as many people as possible [15].

Based on process evaluations of the original randomized controlled trial [12] and a previous public dissemination [19], Germ Defence has been updated and streamlined for use during the COVID-19 outbreak, including broadening the infection control behaviors that were recommended. The intervention is a single session designed to be easily accessible with no sign-up or password required. Full details of the intervention structure and development are reported elsewhere [3,16,18,19] and archived copies are available at [20] (see Germ Defence v3). Intervention content was "frozen" during the reported data collection period. A structured outline of content is available in Textbox 1.

Textbox 1. A detailed outline of Germ Defence content and structure (note: the website and all associated content can be accessed for free).

Introductory content (3 pages)

- Introductory pages seek to increase users' perceived risk by emphasizing the personal and social health consequences of contracting COVID-19. These are followed by messages to increase skills and confidence to reduce exposure to the virus.

Website pathway selection (2 pages)

- To allow users to choose the advice they consider most personally relevant, the intervention is structured so that users initially select between two components of interest: handwashing and reducing Illness. The reducing illness component is tailored such that a user selects one of four streams of content (each lasting 11 pages) that is relevant to the user's situation: (1) to protect themselves generally, (2) to protect others if the user was showing symptoms, (3) to protect themselves if household member(s) showed symptoms, or (4) to protect a household member who is at high risk. The advice is tailored in this way to encourage users to adopt behaviors appropriate to the perceived level and pattern of risk in their household. For example, users in the protect themselves generally group would vary from very low to very high risk. It was not possible to provide specific tailored advice for every household combination of risks and resources (eg, based on the need and potential for household members to self-isolate within the home); therefore, Germ Defence aimed to educate users to adopt behaviors that were appropriate and feasible for their own circumstances.

Tailored infection control behavior advice (7 pages)

- Clear and detailed advice is then provided for self-isolating, social distancing, disinfecting/cleaning, wearing face coverings, and putting items aside that may have viruses on them such as shopping/packages. Advice is provided to the extent that users feel is appropriate for the perceived risk. These pages also contain ideas and information on how to structure the home and engage in behaviors safely. The handwashing component provides advice focused on handwashing that is relevant to all groups over 5 pages.

Goal-setting advice (3 pages)

- Both the handwashing and reducing illness components contain goal-setting sections where users indicate their behavior over the past week, view a motivational message, and then plan their behavior for the future. Users who do not select any improvement are encouraged to review their plan. After completing either the handwashing or reducing illness components, users are asked how helpful they found the website.

Additional information

- Users are then able to revisit the first two components, choose from two additional components with more detailed information about the same behaviors (eg, how to social distance with young children, how to stop touching your face), or view details about the website.

Statistical Analysis

We included data from all users who accessed the website during the study period.

For analysis, users were grouped according to the tailored website pathway they selected within the *reducing illness* component (*protect myself generally* vs *protect others if I am showing symptoms* vs *protect myself if a household member has symptoms* vs *protect a household member at high risk*). Users could also view the handwashing component, which was relevant to all groups. If they did not view *reducing illness*, they were not included in group comparisons, but handwashing responses were still recorded. Users could complete more than one type of tailored pathway, but we only analyzed responses for the pathway that was selected first.

To understand current infection control behaviors (aim 1), behavioral measures were analyzed individually and collapsed together to form an *average infection control behavior* score. When users completed a plan more than once (eg, if they received website feedback that their initial plan could be further improved), the *final* plan was used. If users did not think a behavior was relevant to them (eg, they lived alone so did not need to socially isolate or could not socially isolate from young children), they could answer *not applicable*. This was coded as missing data and not included in analysis. Linear regression compared between-group scores for behavior.

To compare current behaviors with intended behavior after using Germ Defence (aim 2), linear regression models comparing between-group scores for intentions controlled for current behavior were used. Paired *t* test comparisons examined the

difference between current behavior and intended behavior within groups.

Results

Use of the Germ Defence Website

We considered data from 53,125 users who completed at least the initial introductory website pages. Users accessed Germ Defence from 129 countries (a full CONSORT diagram of use is presented in Figure 1). The majority (n=44,446, 83.7%) of users were from the United Kingdom (England: n=40,164, 75.6%; Scotland: n=2204, 4.2%; Wales: n=1459, 2.8%; Northern Ireland: n=566, 1.1%; other: n=73, 0.1%). The mean use time was 8 minutes 28 seconds, and the mean number of pages viewed was 19.9. Of the recorded sessions, 54.1% (n=28,740) lasted longer than 1 minute. Over half (n=28,687, 54%) of the users accessed Germ Defence using a mobile device, 31% (n=16,469) accessed with a tablet, and 15% (n=7968) with a desktop or laptop computer. Only 10.6% (n=5631) of users were *return users* visiting for a second time. Aggregated use statistics for users outside the United Kingdom are provided in Multimedia Appendix 1. Detailed use for each website component is presented in Figure 1. The overall mean helpfulness of the website was rated as 7.77 (SD 2.31) out of 10.

Infection Control Behaviors and Intended Behaviors in Users of Germ Defence

All groups (protect themselves generally, protect others if the user was showing symptoms, protect themselves if household members were showing symptoms, and protect a household member who is high risk) reported using most current infection behaviors sometimes or quite often within the home. Overall, users reported they would wear a face covering almost never or sometimes (mean 1.61, SD 1.19) and would socially distance sometimes or quite often (mean 2.40, SD 1.22). Users reported socially isolating in their own room sometimes or quite often (mean 2.78, SD 1.29) and putting packages and shopping aside sometimes or quite often (mean 2.75, SD 1.55). Users reported cleaning and disinfecting quite often or very often (mean 3.17, SD 1.18).

Frequency of the five infection control behaviors from the *reducing illness* pathway within each group is reported in Table 2 (with handwashing reported in a separate table), as well as mean differences and 95% CIs of group comparisons (each group vs the *protect themselves generally* group). The frequency of behaviors did not vary appreciably between groups; numerically, the *protect themselves generally* group were least likely to socially distance (mean 2.39, SD 1.22). People in the *protect others if user showing symptoms* group were least likely to clean and disinfect (mean 2.95, SD 1.26) and put aside shopping and packages (mean 2.39, SD 1.48) but most likely to wear a face covering (mean 1.91, SD 1.36). People in the *protect themselves if household members showing symptoms* group were most likely to maintain social distance (mean 2.57, SD 1.23), and users in the *protect household members at high risk* group were least likely to stay in their own room (mean 2.64, SD 1.16) and least likely to wear a face covering (mean 1.42, SD 0.99).

Table 2 shows some small differences in how often participants planned to perform behaviors in the future (corrected for levels of current behavior) between groups. Compared to people in the protect themselves generally group, people showing symptoms planned to clean and disinfect, and put aside shopping less frequently, but they planned to self-isolate more frequently. People in the protect themselves from household member with symptoms group planned to socially distance and self-isolate more frequently than those in the protect themselves generally group. People looking to protect a high-risk household member planned to conduct all of the behaviors slightly more frequently than the protect themselves generally group.

Paired *t* test comparisons examined differences between current and planned behaviors after using the Germ Defence website. Mean difference scores for each group and 95% CIs are reported in Table 3. The difference between intended and current behavior was largest for cleaning and disinfecting (mean difference 0.38, 95% CI 0.37-0.39) and putting aside shopping and packages (mean difference 0.49, 95% CI 0.47-0.50), and was lowest for self-isolating (mean difference 0.15, 95% CI 0.14-0.16). Overall, infection control behaviors increased (mean difference 0.30, 95% CI 0.29-0.31).

Handwashing behavior is reported in Table 4. Mean current handwashing behavior was higher than other infection control behaviors (mean 4.04, SD 0.84) with reported intended behavior consistently higher (mean increase 0.41, 95% CI 0.40-0.42).

Table 2. Current and intended infection control behaviors.

Behaviors	Protect themselves generally (n=18,029)[a], mean (SD)	Protect others if user showing symptoms (n=169)			Protect themselves if household member showing symptoms (n=319)			Protect a household member at high risk (n=1787)		
		Mean (SD)	Mean difference (95% CI)	Cohen *d*[b]	Mean (SD)	Mean difference (95% CI)	Cohen *d*	Mean (SD)	Mean difference (95% CI)	Cohen *d*
Current behavior										
Social distancing	2.39 (1.22)	2.52 (1.39)	0.13 (–0.07 to 0.33)	0.11	2.57 (1.23)	0.17 (0.04 to 0.31)	0.15	2.51 (1.20)	0.12 (0.06 to 0.18)	0.10
Clean/disinfect	3.18 (1.18)	2.95 (1.26)	–0.24 (–0.42 to –0.06)	0.20	3.05 (1.18)	0.17 (0.04 to 0.31)	0.11	3.19 (1.17)	0.003 (–0.05 to 0.06)	0.00
Put aside shopping/packages	2.74 (1.55)	2.39 (1.48)	–0.35 (–0.60 to –0.11)	0.23	3.00 (1.49)	0.26 (0.08 to 0.44)	0.17	2.82 (1.59)	0.08 (0.004 to 0.16)	0.05
Self-isolate in own room	2.79 (1.30)	2.85 (1.43)	0.05 (–0.15 to 0.25)	0.04	2.75 (1.26)	–0.04 (–0.19 to 0.10)	0.03	2.64 (1.16)	–0.15 (–0.21 to –0.08)	0.11
Wear face covering	1.63 (1.21)	1.91 (1.36)	0.28 (0.07 to 0.49)	0.24	1.75 (1.28)	0.12 (–0.02 to 0.27)	0.10	1.42 (0.99)	–0.21 (–0.27 to –0.14)	0.17
Overall behavior score[c]	2.67 (0.91)	2.61 (1.08)	–0.05 (–0.19 to 0.08)	0.06	2.68 (0.90)	0.01 (–0.09 to 0.11)	0.01	2.59 (0.80)	–0.07 (–0.12 to –0.03)	0.08
Intended Behavior										
Social distancing	2.63 (1.28)	2.79 (1.47)	0.05 (–0.06 to 0.16)[d]	0.12	2.88 (1.30)	0.12 (0.05 to 0.20)[d]	0.19	2.84 (1.27)	0.11 (0.07 to 0.14)[d]	0.16
Clean/disinfect	3.57 (1.16)	3.18 (1.33)	–0.14 (–0.25 to –0.03)[d]	0.33	3.46 (1.18)	0.001 (–0.08 to 0.08)[d]	0.09	3.63 (1.15)	0.05 (0.01 to 0.08)[d]	0.05
Put aside shopping/packages	3.24 (1.52)	2.73 (1.59)	–0.19 (–0.34 to –0.04)[d]	0.34	3.44 (1.41)	–0.02 (–0.12 to 0.09)[d]	0.13	3.37 (1.52)	0.06 (0.01 to 0.11)[d]	0.08
Self-isolate in own room	2.94 (1.28)	3.08 (1.41)	0.10 (0.02 to 0.18)[d]	0.12	2.97 (1.23)	0.07 (0.01 to 0.13)[d]	0.03	2.87 (1.17)	0.06 (0.04 to 0.09)[d]	0.05
Wear face covering	1.95 (1.37)	2.19 (1.50)	0.03 (–0.11 to 0.17)[d]	0.18	2.15 (1.47)	0.08 (–0.01 to 0.18)[d]	0.15	1.82 (1.28)	0.08 (0.03 to 0.12)[d]	0.09
Overall behavior score	2.97 (0.96)	2.86 (1.20)	–0.03 (–0.12 to 0.05)[d]	0.11	3.01 (0.96)	0.03 (–0.03 to 0.09)[d]	0.04	2.97 (0.89)	0.06 (0.03 to 0.08)[d]	0.00

[a]Between group comparisons compare each group to the protect themselves generally group. Scale: 1 is almost never, 2 is sometimes, 3 is quite often, 4 is very often, and 5 is almost always.

[b]Reported as the standardized mean difference between each group and the comparison group.

[c]Overall behavior scores are means calculated from all behaviors in which a response was recorded.

[d]Controlling for current behavior.

Table 3. Group differences between behavior and intention.

Behaviors	Protect themselves generally (n=18,029)[a]		Protect others if user showing symptoms (n=169)		Protect themselves if household member showing symptoms (n=319)		Protect a household member at high risk (n=1787)		Overall	
	Mean difference (95% CI)	Cohen *d*	Mean difference (95% CI)	Cohen *d*	Mean difference (95% CI)	Cohen *d*	Mean difference (95% CI)	Cohen *d*	Mean difference (95% CI)	Cohen *d*
Behavior										
Social distancing	0.22 (0.21-0.23)	0.35	0.26 (0.11-0.40)	0.30	0.33 (0.24-0.42)	0.41	0.31 (0.28-0.35)	0.43	0.23 (0.22-0.24)	0.36
Clean/disinfect	0.38 (0.37-0.39)	0.52	0.30 (0.17-0.44)	0.36	0.41 (0.31-0.51)	0.47	0.43 (0.39-0.47)	0.54	0.38 (0.37-0.40)	0.52
Put aside shopping/packages	0.49 (0.47-0.50)	0.49	0.39 (0.24-0.54)	0.42	0.41 (0.31-0.51)	0.47	0.53 (0.48-0.58)	0.50	0.49 (0.47-0.50)	0.49
Self-isolate in own room	0.14 (0.13-0.15)	0.28	0.23 (0.11-0.36)	0.30	0.21 (0.14-0.29)	0.33	0.22 (0.19-0.25)	0.34	0.15 (0.14-0.16)	0.29
Wear face covering	0.28 (0.27-0.30)	0.37	0.29 (0.12-0.47)	0.30	0.35 (0.25-0.46)	0.42	0.37 (0.33-0.42)	0.42	0.29 (0.28-0.29)	0.37
Average infection control score	0.29 (0.29-0.30)	0.53	0.27 (0.16-0.38)	0.38	0.32 (0.25-0.40)	0.49	0.36 (0.33-0.39)	0.57	0.30 (0.29-0.31)	0.53

[a]Group n values are taken across all behaviors.

Table 4. Paired comparisons between current and intended handwashing behavior.

Handwashing situation	Current behavior (n=12,981), mean (SD)	Intended behavior (n=12,981), mean (SD)	Mean difference (95% CI)	Cohen *d*
Before eating snacks	3.91 (1.28)	4.45 (0.99)	0.54 (0.52-0.56)	0.54
After coming home	4.66 (0.81)	4.80 (0.62)	0.14 (0.13-0.15)	0.26
After sneezing or coughing	3.45 (1.43)	4.11 (1.23)	0.66 (0.64-0.68)	0.59
After contact with possible carrier	4.22 (1.24)	4.53 (1.00)	0.30 (0.29-0.32)	0.36
After touching something	4.13 (1.23)	4.50 (0.97)	0.36 (0.35-0.38)	0.43
Overall score[a]	4.00 (1.03)	4.34 (0.91)	0.34 (0.33-0.35)	0.50

[a]Handwashing overall score was a separate item

Discussion

Summary of Findings

Germ Defence was accessed by a large number of users across 129 countries, primarily from the United Kingdom. This demonstrates public interest in adopting appropriate infection control behaviors in the home during the COVID-19 pandemic. After using Germ Defence, all groups reported intentions to increase the frequency of their infection control behaviors, including handwashing.

Except for handwashing, self-reported infection control behaviors in the home were only reported *sometimes or quite often* regardless of whether people were seeking to protect themselves, concerned about demonstrating COVID-19 symptoms, had a household member showing symptoms, or were seeking to protect a high-risk household member. The frequency of wearing face coverings was consistently the lowest of the behaviors, while cleaning and disinfecting was the most frequently reported of the behaviors outside of handwashing. All of these infection control behaviors were reported to be performed much less frequently than handwashing.

As would be expected, certain behaviors and intentions varied according to the circumstances of groups; for example, people seeking to protect others when showing symptoms reported higher current frequencies of wearing face coverings, while people seeking to protect a high-risk household member reported the intention to socially distance within the home more frequently.

Comparison With Existing Literature

This study provides the first up-to-date analysis of infection control behaviors and intentions across the United Kingdom in a large sample during the COVID-19 pandemic. Within-household transmission will be increasingly important as infection control measures become established in external, public environments [6,21]. Therefore, understanding current

infection control behaviors within homes (and how to improve them) is vital to continue controlling the pandemic.

Self-reported infection control behaviors other than handwashing are lower than is optimal for infection prevention, even in Germ Defence users who were likely more motivated and willing to engage in protective behaviors than the general population (as they were seeking additional information) [22]. Increasing engagement in these behaviors is important as societal restrictions are released and perceived risk reduces [23].

Germ Defence users reported intentions to increase the frequency of infection control behaviors over their current rates. Although such intentions potentially misrepresent the observed behavioral change after an intervention (the *intention-behavior gap* [24]), our evidence suggests that Germ Defence may overcome this. Analysis of comparable data from the PRIMIT trial handwashing intervention showed slightly smaller behavior and intention differences (Cohen d=0.45). This change was sufficient to cause reduced infection transmission and severity within households after 16 weeks [12]. Comparable data during the current pandemic (reducing illness behaviors: Cohen d=0.53; handwashing: Cohen d=0.50) shows a slightly larger effect across a broader range of behaviors that may have a larger impact on infection rates.

Study Limitations

As a cross-sectional observation of an active intervention, Germ Defence lacks longitudinal follow-up. Care must be taken when interpreting findings within the rapidly changing context of the COVID-19 pandemic. Our method of categorization using website pathways may not be accurate for some users or might overlook individual differences within categories.

Our data may not be a representative sample from the wider UK population for several reasons. First, users of Germ Defence are likely to be more motivated and report higher frequencies of infection control behaviors. Second, although analytic data indicates that the large majority of the intervention's users were from the United Kingdom, we could not identify non-UK users within behavioral data. Finally, self-reported infection control behaviors may not be accurate reflections of actual behaviors occurring within households.

However, none of these limitations affect our main findings; indeed, people are prone to overreport protective behaviors, further highlighting the need for improvement.

Implications for Practice and Research

A concerted effort to improve household infection control behaviors across the UK population is likely to be an efficient use of health resources, both to reduce current rates of infection and to prevent the likelihood and severity of future outbreaks. Handwashing behaviors are already relatively high—perhaps due to existing familiarity with the behavior supported by a focus in public health advice on increasing handwashing in earlier stages of the pandemic. Therefore recommending digital interventions such as Germ Defence to target other infection control behaviors within the home may help control the current pandemic.

Given the current rates of infection control behaviors within the home even within a motivated sample, it is vital to address barriers to engaging in them. For example, people living in crowded, working households are more likely to come into contact with the virus [5] and may find it difficult to self-isolate. Similarly, cultural differences, financial challenges, or caring responsibilities may cause barriers to social distancing [6]. Research should explore how to support these behaviors for as many households as possible. Indeed, digital interventions such as Germ Defence can use tailored content to target behaviors that are relevant for specific user groups.

Conclusion

Our findings show substantial room for improvement in protective behaviors across the United Kingdom—even in our motivated, self-selected sample—as societal restrictions are eased. People are not sufficiently self-isolating within the home to prevent household transmission, even when a household member or the individual themselves are demonstrating COVID-19 symptoms. Promoting evidence-based behavior change interventions might improve these behaviors, reducing transmission within households and the incidence and severity of infections.

Germ Defence is a scalable, evidence-based, acceptable, and free public health intervention with negligible safety risk, which could be included in public heath guidance and promoted via primary care networks at minimal cost for wide population coverage.

Acknowledgments

We thank all who have supported and disseminated Germ Defence, including the University of Bath (Andy Dunne) and the National Institute for Health Research (NIHR) Bristol Health Protection Research Unit (Helen Bolton and Clare Thomas). We thank all who have assisted in the translation of Germ Defence into other languages [9]. We also thank the many citizen scientists and public contributors who assisted in the development of the Germ Defence intervention.

The study was funded by the United Kingdom Research and Innovation (UKRI)/Medical Research Council Rapid Response Call: UKRI CV220-009.

The Germ Defence intervention was hosted by the Lifeguide Team, supported by the NIHR Biomedical Research Centre, University of Southampton. LY is a NIHR Senior Investigator and theme lead for University of Southampton Biomedical Research Centre. LY and RA are affiliated to the NIHR Health Research Health Protection Research Unit in Behavioural Science and Evaluation of Interventions at the University of Bristol in partnership with Public Health England. MW is an NIHR Academic Clinical

Lecturer, under grant CL-2016-26-005. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, the Department of Health, or Public Health England. The funders had no role in the design of the study, collection, analysis, and interpretation of data or in writing the manuscript.

Authors' Contributions

BA, SM, and LY conceived the study. BA and LY developed the study design. BA, BS, and JG analyzed the design. All authors interpreted the data and developed the intervention. BA drafted the manuscript. All authors reviewed the manuscript and approved the content.

BA and LY confirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and any discrepancies have been explained.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison of aggregated use statistics for users outside of the United Kingdom (compared to users within the United Kingdom).
[DOCX File , 13 KB - jmir_v23i2e22197_app1.docx]

References

1. Cheng H, Jian S, Liu D, Ng T, Huang W, Lin H, Taiwan COVID-19 Outbreak Investigation Team. Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. JAMA Intern Med 2020 Sep 01;180(9):1156-1163 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2020] [Medline: 32356867]
2. Burke RM, Midgley CM, Dratch A, Fenstersheib M, Haupt T, Holshue M, et al. Active monitoring of persons exposed to patients with confirmed COVID-19 - United States, January-February 2020. MMWR Morb Mortal Wkly Rep 2020 Mar 06;69(9):245-246. [doi: 10.15585/mmwr.mm6909e1] [Medline: 32134909]
3. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect Dis 2020 Aug;20(8):911-919 [FREE Full text] [doi: 10.1016/S1473-3099(20)30287-5] [Medline: 32353347]
4. West R, Michie S, Rubin GJ, Amlôt R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. Nat Hum Behav 2020 May;4(5):451-459. [doi: 10.1038/s41562-020-0887-9] [Medline: 32377018]
5. Wang Y, Tian H, Zhang L, Zhang M, Guo D, Wu W, et al. Reduction of secondary transmission of SARS-CoV-2 in households by face mask use, disinfection and social distancing: a cohort study in Beijing, China. BMJ Glob Health 2020 May;5(5):e002794 [FREE Full text] [doi: 10.1136/bmjgh-2020-002794] [Medline: 32467353]
6. Atchison C, Bowman LR, Vrinten C, Redd R, Pristerà P, Eaton J, et al. Early perceptions and behavioural responses during the COVID-19 pandemic: a cross-sectional survey of UK adults. BMJ Open 2021 Jan 04;11(1):e043577 [FREE Full text] [doi: 10.1136/bmjopen-2020-043577] [Medline: 33397669]
7. Band R, Bradbury K, Morton K, May C, Michie S, Mair FS, et al. Intervention planning for a digital intervention for self-management of hypertension: a theory-, evidence- and person-based approach. Implement Sci 2017 Feb 23;12(1):25 [FREE Full text] [doi: 10.1186/s13012-017-0553-4] [Medline: 28231840]
8. Coronavirus (COVID-19): guidance. GOV.UK. 2020. URL: https://www.gov.uk/government/collections/coronavirus-covid-19-list-of-guidance [accessed 2020-06-08]
9. Germ Defence. 2020. URL: https://germdefence.org [accessed 2020-11-23]
10. Gold N, Hu XY, Denford S, Xia RY, Towler L, Groot J, et al. Effectiveness of digital interventions to improve household and community infection prevention and control behaviours and to reduce incidence of respiratory and/or gastro-intestinal infections: a rapid systematic review. medRxiv. 2021 Preprint posted online September 9, 2020. [doi: 10.1101/2020.09.07.20164947]
11. Yardley L, Miller S, Schlotz W, Little P. Evaluation of a web-based intervention to promote hand hygiene: exploratory randomized controlled trial. J Med Internet Res 2011 Dec 09;13(4):e107 [FREE Full text] [doi: 10.2196/jmir.1963] [Medline: 22155673]
12. Little P, Stuart B, Hobbs FDR, Moore M, Barnett J, Popoola D, et al. An internet-delivered handwashing intervention to modify influenza-like illness and respiratory infection transmission (PRIMIT): a primary care randomised trial. Lancet 2015 Oct 24;386(10004):1631-1639. [doi: 10.1016/S0140-6736(15)60127-1] [Medline: 26256072]
13. Germ Defence JMIR - Ainsworth et al 2021. Figshare. URL: https://figshare.com/s/72e9fbfd1f7bbf090220 [accessed 2021-02-19]
14. Yardley L, Miller S, Teasdale E, Little P, Primit Team. Using mixed methods to design a web-based behavioural intervention to reduce transmission of colds and flu. J Health Psychol 2011 Mar;16(2):353-364. [doi: 10.1177/1359105310377538] [Medline: 20929941]

15. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
16. Ajzen I. The theory of planned behavior. Organizational Behav Hum Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
17. Leventhal H, Phillips LA, Burns E. The Common-Sense Model of Self-Regulation (CSM): a dynamic framework for understanding illness self-management. J Behav Med 2016 Dec;39(6):935-946. [doi: 10.1007/s10865-016-9782-2] [Medline: 27515801]
18. Rippetoe PA, Rogers RW. Effects of components of protection-motivation theory on adaptive and maladaptive coping with a health threat. J Pers Soc Psychol 1987 Mar;52(3):596-604. [doi: 10.1037//0022-3514.52.3.596] [Medline: 3572727]
19. Ainsworth B, Steele M, Stuart B, Joseph J, Miller S, Morrison L, et al. Using an analysis of behavior change to inform effective digital intervention design: how did the PRIMIT website change hand hygiene behavior across 8993 users? Ann Behav Med 2017 Jun;51(3):423-431 [FREE Full text] [doi: 10.1007/s12160-016-9866-9] [Medline: 27909944]
20. Germ Defense archive. URL: http://archive.germdefence.org/ [accessed 2021-02-19]
21. Little P, Read RC, Amlôt R, Chadborn T, Rice C, Bostock J, et al. Reducing risks from coronavirus transmission in the home-the role of viral load. BMJ 2020 May 06;369:m1728. [doi: 10.1136/bmj.m1728] [Medline: 32376669]
22. Jenner EA, Fletcher BC, Watson P, Jones FA, Miller L, Scott GM. Discrepancy between self-reported and observed hand hygiene behaviour in healthcare professionals. J Hosp Infect 2006 Aug;63(4):418-422. [doi: 10.1016/j.jhin.2006.03.012] [Medline: 16772101]
23. Dryhurst S, Schneider C, Kerr J, Freeman A, Recchia G, van der Bles AM, et al. Risk perceptions of COVID-19 around the world. J Risk Res 2020 May 05;23(7-8):994-1006. [doi: 10.1080/13669877.2020.1758193]
24. Sniehotta F, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychol Health 2005 Apr;20(2):143-160. [doi: 10.1080/08870440512331317670]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
NIHR: National Institute for Health Research
PRIMIT: Primary Care Randomised Trial of an Internet Intervention to Modify Influenza-Like Illness and Respiratory Infection Transmission
UKRI: United Kingdom Research and Innovation

Edited by G Eysenbach; submitted 06.07.20; peer-reviewed by D Baretta, H Akram, H Yasuda; comments to author 17.11.20; revised version received 15.12.20; accepted 24.01.21; published 25.02.21.

Please cite as:
Ainsworth B, Miller S, Denison-Day J, Stuart B, Groot J, Rice C, Bostock J, Hu XY, Morton K, Towler L, Moore M, Willcox M, Chadborn T, Gold N, Amlôt R, Little P, Yardley L
Infection Control Behavior at Home During the COVID-19 Pandemic: Observational Study of a Web-Based Behavioral Intervention (Germ Defence)
J Med Internet Res 2021;23(2):e22197
URL: https://www.jmir.org/2021/2/e22197
doi:10.2196/22197
PMID:33566791

Original Paper

Using Automated Machine Learning to Predict the Mortality of Patients With COVID-19: Prediction Model Development Study

Kenji Ikemura[1,2], MD; Eran Bellin[3], MD; Yukako Yagi[4], PhD; Henny Billett[5], MD; Mahmoud Saada[2], BS; Katelyn Simone[2], MM; Lindsay Stahl[3], MS; James Szymanski[1], MD; D Y Goldstein[1], MD; Morayma Reyes Gil[1], MD, PhD

[1]Department of Pathology, Albert Einstein College of Medicine, Montefiore Medical Center, The Bronx, NY, United States

[2]Tsubomi Technology, The Bronx, NY, United States

[3]Department of Epidemiology and Population Health and Medicine, Albert Einstein College of Medicine, Montefiore Medical Center, The Bronx, NY, United States

[4]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, United States

[5]Department of Oncology and Medicine, Albert Einstein College of Medicine, Montefiore Medical Center, The Bronx, NY, United States

Corresponding Author:
Kenji Ikemura, MD
Department of Pathology
Albert Einstein College of Medicine
Montefiore Medical Center
111 E 210th St
The Bronx, NY, 10467
United States
Phone: 1 9493703777
Email: kikemura@montefiore.org

Abstract

Background: During a pandemic, it is important for clinicians to stratify patients and decide who receives limited medical resources. Machine learning models have been proposed to accurately predict COVID-19 disease severity. Previous studies have typically tested only one machine learning algorithm and limited performance evaluation to area under the curve analysis. To obtain the best results possible, it may be important to test different machine learning algorithms to find the best prediction model.

Objective: In this study, we aimed to use automated machine learning (autoML) to train various machine learning algorithms. We selected the model that best predicted patients' chances of surviving a SARS-CoV-2 infection. In addition, we identified which variables (ie, vital signs, biomarkers, comorbidities, etc) were the most influential in generating an accurate model.

Methods: Data were retrospectively collected from all patients who tested positive for COVID-19 at our institution between March 1 and July 3, 2020. We collected 48 variables from each patient within 36 hours before or after the index time (ie, real-time polymerase chain reaction positivity). Patients were followed for 30 days or until death. Patients' data were used to build 20 machine learning models with various algorithms via autoML. The performance of machine learning models was measured by analyzing the area under the precision-recall curve (AUPCR). Subsequently, we established model interpretability via Shapley additive explanation and partial dependence plots to identify and rank variables that drove model predictions. Afterward, we conducted dimensionality reduction to extract the 10 most influential variables. AutoML models were retrained by only using these 10 variables, and the output models were evaluated against the model that used 48 variables.

Results: Data from 4313 patients were used to develop the models. The best model that was generated by using autoML and 48 variables was the stacked ensemble model (AUPRC=0.807). The two best independent models were the gradient boost machine and extreme gradient boost models, which had an AUPRC of 0.803 and 0.793, respectively. The deep learning model (AUPRC=0.73) was substantially inferior to the other models. The 10 most influential variables for generating high-performing models were systolic and diastolic blood pressure, age, pulse oximetry level, blood urea nitrogen level, lactate dehydrogenase level, D-dimer level, troponin level, respiratory rate, and Charlson comorbidity score. After the autoML models were retrained with these 10 variables, the stacked ensemble model still had the best performance (AUPRC=0.791).

Conclusions: We used autoML to develop high-performing models that predicted the survival of patients with COVID-19. In addition, we identified important variables that correlated with mortality. This is proof of concept that autoML is an efficient, effective, and informative method for generating machine learning–based clinical decision support tools.

(J Med Internet Res 2021;23(2):e23458) doi:10.2196/23458

KEYWORDS

automated machine learning; COVID-19; biomarker; ranking; decision support tool; machine learning; decision support; Shapley additive explanation; partial dependence plot; dimensionality reduction

Introduction

Many regions worldwide are still fighting the first wave of the COVID-19 pandemic, while other areas that have reopened are experiencing a resurgence of cases [1]. During such an emergent situation, it is important for clinicians to effectively and efficiently triage patients. In recent months, studies have proposed several machine learning models that can accurately predict COVID-19 disease severity. Many of these studies have been successful in generating a high-performing model [2-4]. However, until now, these models have only been trained on one kind of machine learning algorithm, and many researchers have limited the evaluation of their models' performance to area under the curve (AUC) analysis. Studies have either not reported areas under the precision-recall curve (AUPRCs) or have only reported low AUPRCs. Furthermore, these studies have been difficult to replicate due to hyperparameter tuning. The automation of machine learning end-to-end processes has allowed for the development of simple, fast, and easy-to-replicate models that often outperform manually designed models. This study was designed to (1) optimize the performance of predictive models by using automated machine learning (autoML) to generate various machine learning models and automate hyperparameter optimization; and (2) choose the best performing machine learning model based on AUPRCs.

Artificial intelligence (ie, machine learning) models have often been criticized for being black-box models. We tried to stare into this so-called "black box," identify the variables that drive model performance, and understand the extent of these variables' effects on model performance. The interpretability of models is crucial in medical environments; for results to be widely accepted, they must be explainable to medical providers. To assess the correctness of a model, clinicians must be able to use their intuition. Therefore, a model's response must be understandable to clinicians and comparable to biologically plausible expectations.

In this study, we aimed to generate multiple machine learning models, assess their performance, and select the highest-performing model. After ranking variables by importance, we chose the top 10 most influential variables and retrained the autoML models to generate new models that only used these 10 variables. This was done to create high-performing models with low dimensionality. In addition, we sought to provide interpretable black-box model results to clinicians and patients. Finally, the COVID-19 mortality calculator, which is based on this study, was developed and freely available online as a web application [5].This study provides proof of concept that autoML is an efficient, effective, and informative method for building machine learning–based clinical decision support tools.

Methods

Variable Selection and Collection

After conducting a literature review, we selected 48 variables for generating high-performing machine learning models. These variables included demographics such as gender; race; age; comorbidities; physical signs/symptoms; and laboratory test results, such as ferritin, interleukin-6, tumor necrosis factor-α, D-dimer, C-reactive protein, and lactic dehydrogenase (LDH) levels [2-4, 6-12].

Data collection and analysis were approved by the Albert Einstein College of Medicine Institutional Review Board. The data were collected by using Clinical Looking Glass (CLG), which is an interactive software application that was developed at the Montefiore Medical Center. This application is used to evaluate health care quality, effectiveness, and efficiency. CLG integrates clinical and administrative data sets, thereby allowing clinicians to build temporally sophisticated cohorts and assess outcomes [13-16].

We queried the CLG database for patients who were aged >18 years, tested positive for COVID-19 (ie, confirmed with a nasopharyngeal specimen and real-time polymerase chain reaction) within 24 hours before or after admission, and were admitted to our institution from March 1 to July 3, 2020. The index time was when a patient tested positive for COVID-19 based on their real-time polymerase chain reaction results. We investigated a total of 48 variables and used the earliest values that were available within 36 hours before or after the index time. The outcome of interest was mortality from any cause within 30 days after the index time.

Model Development and Evaluation

We used the open-source H2O.ai autoML package for the R language [17-19]. The package can be downloaded to a local device. This allowed us to avoid uploading patient data to a third-party cloud service. The H2O.ai autoML package trains and cross-validates common machine learning algorithms, such as gradient boosting machine (GBM), extreme gradient boosting (XGBoost), general linear models (GLMs), random forest (RF), and deep learning (DL). In addition, the package trains two types of stacked ensemble models—one based on all previously trained models and another based on the best model of each model family. Additional information on how each model was built and which hyperparameters were tuned via autoML can be found in documentation that is provided by H2O.ai [18,19].

We used the 10-fold cross-validation method to train the autoML model on a randomly selected data set that included 80% of the original data. We then used the trained autoML model to generate 20 models and rank them in order of performance (ie, AUPRC). These 20 models were based on the remaining 20% of the original data set.

The AUPRC is a measure of a model's predictive performance, which is based on the relationship between the positive predictive value (PPV) for the outcome (ie, death; y-axis) and the model's sensitivity for detecting patients who actually die (ie, x-axis). For reproducibility, we did not use the DL method and trained each model separately. For convenience, we named the best model that was generated with 48 variables MODEL-48. After creating Shapley additive explanation (SHAP) and partial dependence (PD) plots to evaluate MODEL-48, we selected the 10 most influential variables. We used these 10 variables to repeat model training, model ranking, and the selection of the best performing autoML-generated model. For convenience, we named the best model that was generated with these 10 variables MODEL-10.

To further evaluate MODEL-48 and MODEL-10, we generated a binary classifier (ie, dead or alive within 30 days). We chose a threshold probability that maximized the F2 score of each model. Unlike the F1 score, which gives equal weight to precision (ie, PPV) and sensitivity (ie, recall), the F2 score gives more weight to sensitivity and penalizes a model for generating more false negatives than false positives. As our goal was to identify patients who were at a high risk of death (ie, patients who need more attention and further intervention), our model's metric of success was based on enhancing its sensitivity for detecting patients who were at risk of death. However, this came with the drawback of overcalling death as a predicted outcome. Sensitivity, specificity, PPVs, and negative predicative values (NPVs) were calculated for each binary classifier. The F-score calculation formula was as follows:

In this formula, to calculate the F2 score, β must equal 2.

Opening the Black Box: Intuitive Understanding of Model Variable Utility

Once a model determined the most important variable in its internal black box, we used SHAP and PD plots to develop our understanding of the black box. A SHAP plot displays variables in a top-to-bottom format; the most important variable is displayed at the top and the least important variable is displayed at the bottom. Variable importance is determined by the model in question. In this study, SHAP values (ie, x-axis) were indicative of the relative contribution of each patient's variable values (eg, a systolic blood pressure of 50 mmHg) to the overall prediction of a patient's mortality. SHAP values of >0 on the x-axis were indicative of variables that contribute to a greater chance of mortality, and SHAP values of <0 were indicative of variables that contribute to a lower chance of mortality. In our SHAP plots, each patient was represented by a dot on a horizontal line (ie, a line for each variable). Each dot's color reflected patients' variable values, which were scaled to a normal, color-coded distribution (ie, red indicates large values and blue indicates small values) [20].

A PD plot is a graphical depiction of a variable's marginal effect on the predicted outcome (ie, mortality). The effect of a variable was measured with mean response values. In this study, mortality had a response value of 1, which indicates a 100% chance of dying. A PD plot can show whether the relationship between a target and a feature is linear, monotonic, or complex [21].

Choosing the Top 10 Most Important Variables: Dimensionality Reduction

Dimensionality reduction is an important process in machine learning model development. Sometimes, the variables in a model correlate with each other, making them redundant variables (ie, blood urea nitrogen [BUN] level, creatinine level, and estimated glomerular filtration rate are all indicators of renal function). If we could generate a model with a low number of unique variables, we would be able to shorten computation times in real clinical settings. In addition, dimension reduction allows models to overcome data sparsity by using variables that have more data points. Furthermore, by identifying the top 10 most important variables, clinicians can focus on ordering medical tests instead of obtaining data on 48 variables, and machine learning developers can have fewer concerns about handling missing values.

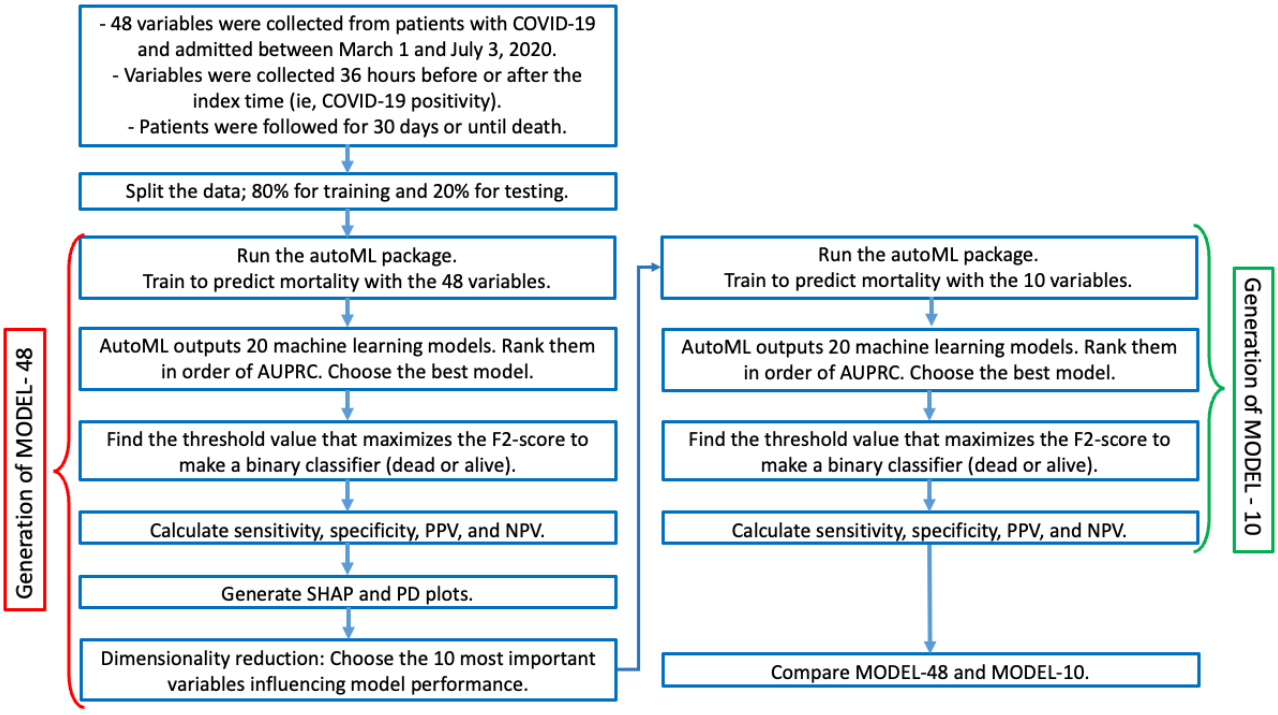
In this study, after evaluating SHAP and PD plots, we chose the 10 most influential variables for generating MODEL-10 (ie, a model that requires only 10 input variables to predict mortality). We first ranked each variable's influence according to the SHAP values in the highest-performing models. Afterward, we chose variables that were influential in these models. Subsequently, if the rank of a variable was not the same in each model, we chose variables based on clinical insights. Clinically speaking, we wanted at least one unique variable for each biological process (ie, cardiac processes, renal processes, coagulation processes, etc). If there was more than a single variable for describing the same clinical domain or biological process (ie, troponin and probrain natriuretic peptide levels), we chose the variable with the fewest number of missing data points (ie, variables that are commonly ordered by clinicians).

Handling Missing Values

Different autoML models have different methods for handling missing values. For example, in a tree-based model (ie, GBM, XGBoost, RF models), missing values are interpreted as data that contain information (ie, data that are missing for a reason) instead of data that are missing at random. During tree building, split decisions are made at every node. Each decision is based on the option that minimizes the loss of model functionality and treats missing values as a separate category. This category is then used as the basis for another split decision. Alternatively, GLMs and DL models use the mean imputation method to handle missing values. Further explanations for how each model imputes missing values can be found in documentation that is provided by H2O.ai [18].

The workflow of our study design is depicted in Figure 1.

Figure 1. Flowchart summary of our methodology. AutoML: automated machine learning; AUPRC: area under the precision-recall curve; NPV: negative predictive value; PD: partial dependence; PPV: positive predictive value; SHAP: Shapley additive explanation.



Data Access, Responsibility, and Analysis

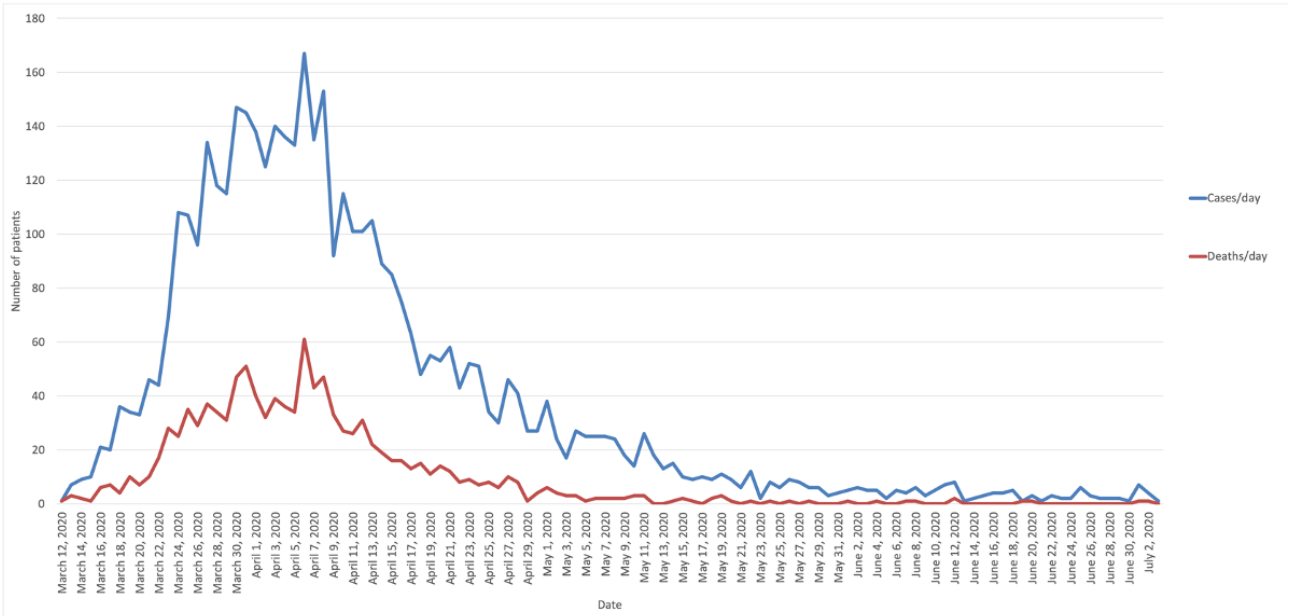
KI had full access to all the data in this study. KI takes responsibility for the integrity of the data and the accuracy of the data analysis.

Results

Study Population

Between March 1 and July 3, 2020, 4313 adult patients tested positive for COVID-19 and were admitted to a Montefiore Health System hospital within 24 hours of their first COVID-19–positive test. Of these 4313 patients, 1087 (25.2%) died within 30 days of infection (Figure 2).

Figure 2. A graph that shows the number of patients who were admitted to the hospital due to SARS-CoV-2 infection (ie, blue line; number of cases per day) and the number of patients who died (ie, red line; number of deaths per day). Data were collected from March 1 to July 3, 2020 at the Montefiore Medical Center.



A summary of case data, patients' survival rates, and patients' demographic characteristics is shown in Table 1. The training set consisted of 3468 patients, and the test set consisted of 845 patients. Summaries of the variables for the entire cohort, the training data set, and the testing data sets are provided in Tables S1, S2, and S3 in Multimedia Appendix 1.

Table 1. Summary of patients' demographic characteristics.

Characteristics	Value
Age (years), mean (SD)	63.97 (16.77)
Gender, n (%)	
Male	2289 (53.07)
Female	2024 (46.93)
Hispanic ethnicity, n (%)	1785 (41.38)
Race, n (%)	
Asian	113 (2.61)
Black	1560 (36.17)
White	428 (9.92)
Other Pacific Islander	4 (0.09)
Native American or Alaskan	5 (0.12)
Unknown/undeclared	418 (9.69)
Other	1785 (41.39)
Charlson score, mean (SD)	2.3 (2.34)
Charlson score of 1, n (%)	
Myocardial infarction	235 (5.44)
Congestive heart failure	702 (16.28)
Peripheral vascular disease	145 (3.36)
Cerebrovascular disease	334 (7.74)
Dementia	716 (16.6)
Chronic pulmonary disease	1030 (23.88)
Rheumatic disease	70 (1.62)
Peptic ulcer disease	33 (0.77)
Mild liver disease	197 (4.57)
Diabetes without chronic complications	616 (14.28)
Charlson score of 2, n (%)	
Diabetes with chronic complications	999 (23.16)
Hemiplegia or paraplegia	118 (2.74)
Renal disease	1314 (30.47)
Any malignancy	178 (4.13)
Charlson score of 3, n (%)	
Moderate or severe liver disease	28 (0.65)
Charlson score of 6, n (%)	
Metastatic solid tumor	58 (1.34)
AIDS/HIV	40 (0.93)
Survivors after 30 days, n (%)	3226 (74.8)
Length of hospital stay to death (number of days)[a], mean (SD)	8.4 (6.91)

[a]In total, 1087 patients died within 30 days of infection.

MODEL-48 Generation and Performance

The output of the 20 machine learning models that were trained via autoML is depicted in Table 2. The best performing model was the stacked ensemble of all machine learning models (AUPRC=0.806). This was MODEL-48. The best performing independent models were the GBM and XGBoost models, which had an AUPRC of 0.803 and 0.793, respectively. The distributed RF model (AUPRC=0.783) came in 14th place and the GLM (AUPRC=0.738) came in last place. The DL model was generated separately from the autoML models for reproducibility purposes. The AUPRC of the DL model plateaued at 0.736.

Table 2. Output of the automated machine learning models that used 48 variables. Model ranks are ordered according to AUPRCs[a].

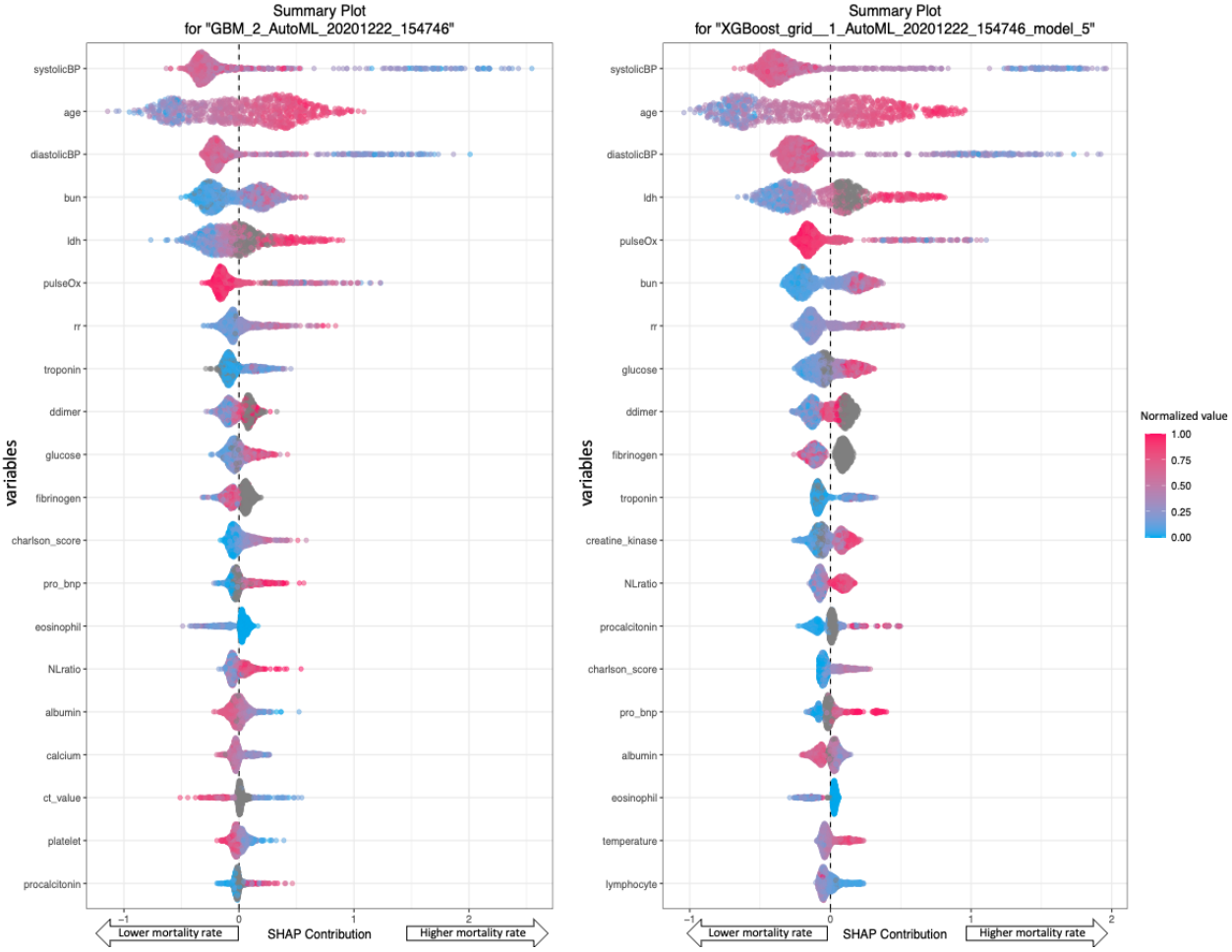
Rank	Model ID	AUPRC	Area under the curve
1	StackedEnsemble_AllModels_AutoML_20201219_141057	0.807	0.917
2	GBM_2_AutoML_20201219_141057	0.803	0.911
3	StackedEnsemble_BestOfFamily_AutoML_20201219_141057	0.800	0.912
4	XGBoost_grid__1_AutoML_20201219_141057_model_5	0.793	0.907
5	GBM_5_AutoML_20201219_141057	0.792	0.907
6	GBM_3_AutoML_20201219_141057	0.791	0.908
7	XGBoost_2_AutoML_20201219_141057	0.790	0.905
8	XGBoost_grid__1_AutoML_20201219_141057_model_6	0.790	0.910
9	XGBoost_grid__1_AutoML_20201219_141057_model_4	0.788	0.903
10	XGBoost_3_AutoML_20201219_141057	0.788	0.910
11	GBM_grid__1_AutoML_20201219_141057_model_3	0.785	0.909
12	GBM_grid__1_AutoML_20201219_141057_model_2	0.785	0.898
13	GBM_4_AutoML_20201219_141057	0.784	0.914
14	DRF_1_AutoML_20201219_141057	0.784	0.905
15	GBM_grid__1_AutoML_20201219_141057_model_1	0.782	0.913
16	GBM_1_AutoML_20201219_141057	0.781	0.903
17	XGBoost_grid__1_AutoML_20201219_141057_model_1	0.779	0.896
18	XGBoost_grid__1_AutoML_20201219_141057_model_3	0.779	0.909
19	XRT_1_AutoML_20201219_141057	0.775	0.899
20	XGBoost_grid__1_AutoML_20201219_141057_model_2	0.769	0.893
21	XGBoost_1_AutoML_20201219_141057	0.763	0.899
22	GLM_1_AutoML_20201219_141057	0.738	0.877

[a]AUPRC: area under the precision-recall curve.

Variable Importance

Figure 3 shows the SHAP plots for the GBM and XGBoost models. In these plots, variables were ranked in descending order of importance. Each patient was represented by one dot on each variable line. The horizontal location of each dot indicated whether the effect of a variable was associated with a higher or lower chance of death [22]. Variable-specific SHAP values of >0 indicated an increased risk of death. For example, the GBM and XGBoost models determined that systolic blood pressure was the most important variable, followed by age and diastolic blood pressure.

Figure 3. SHAP summary plots of the GBM and XGBoost models. According to the GBM and XGBoost models, higher systolic blood pressure levels (ie, red dots) were associated with a lower probability of death (ie, the left side of the vertical dotted line), and older age (ie, red dots) was associated with higher probability of death (ie, the right side of the vertical dotted line). BUN: blood urea nitrogen; charlson_score: charlson comorbidity index; ct_value: cycle threshold value; diastolicBP: diastolic blood pressure; GBM: gradient boosting machine; LDH: lactate dehydrogenase; NLratio: neutrophil-lymphocyte ratio; pro_bnp: pro-brain natriuretic peptide; pulseOx: pulse oximetry; rr: respiratory rate; SHAP: Shapley additive explanation; systolicBP: systolic blood pressure; XGBoost: extreme gradient boosting.



PD plots show the marginal effect that one variable can have on the predicted outcome of a machine learning model. PD plots for the most influential variables are depicted in Multimedia Appendix 2. Each line in a PD plot depicts the best performing model in each machine learning algorithm family. For example, in Figure 4, all models determined that percent mortality increased with age (ie, starting at around 50 years of age). Similarly, all models determined that percent mortality increased with glucose level. However, this was only true for glucose levels of <300 mg/dL (Figure 4).

The importance of each variable in every model is represented on a heatmap in Figure 5.

Figure 4. Partial dependence plots for age and glucose level. Partial dependence plots for the other variables are shown in Multimedia Appendix 2. Each line represents a different machine learning algorithm. DRF: distributed random forest; GBM: gradient boosting machine; GLM: generalized linear model; XGBoost, extreme gradient boosting; XRT: extremely randomized trees.

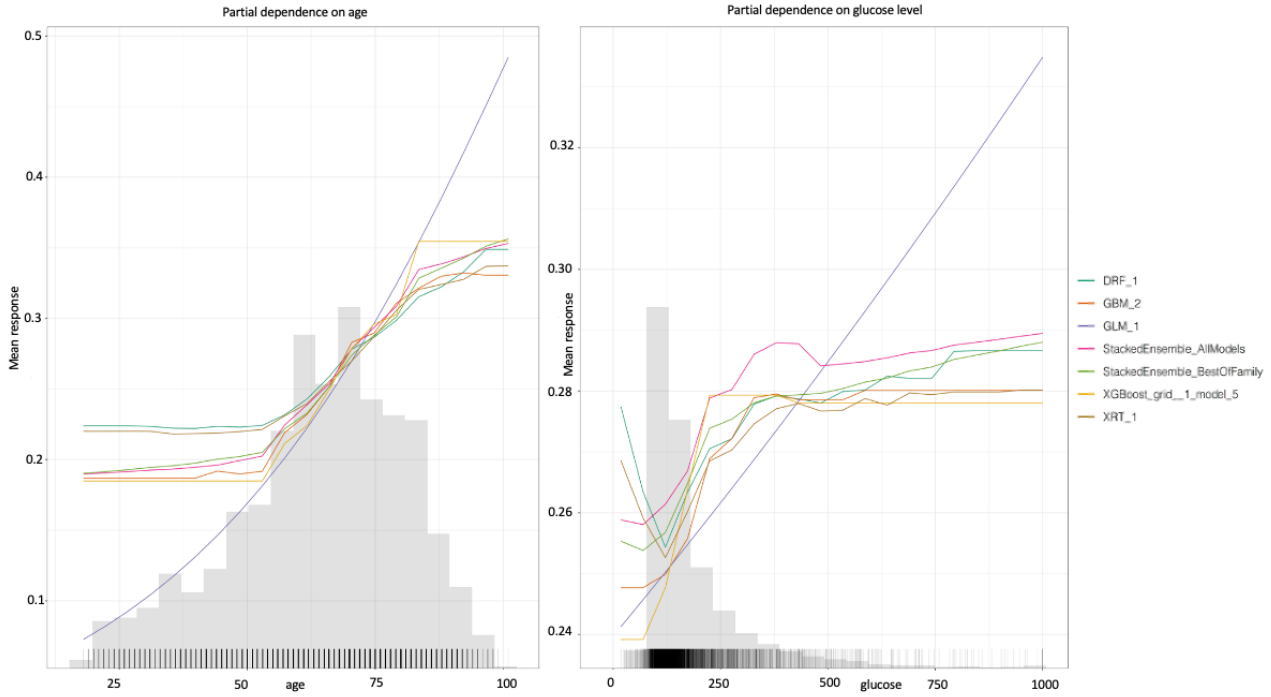
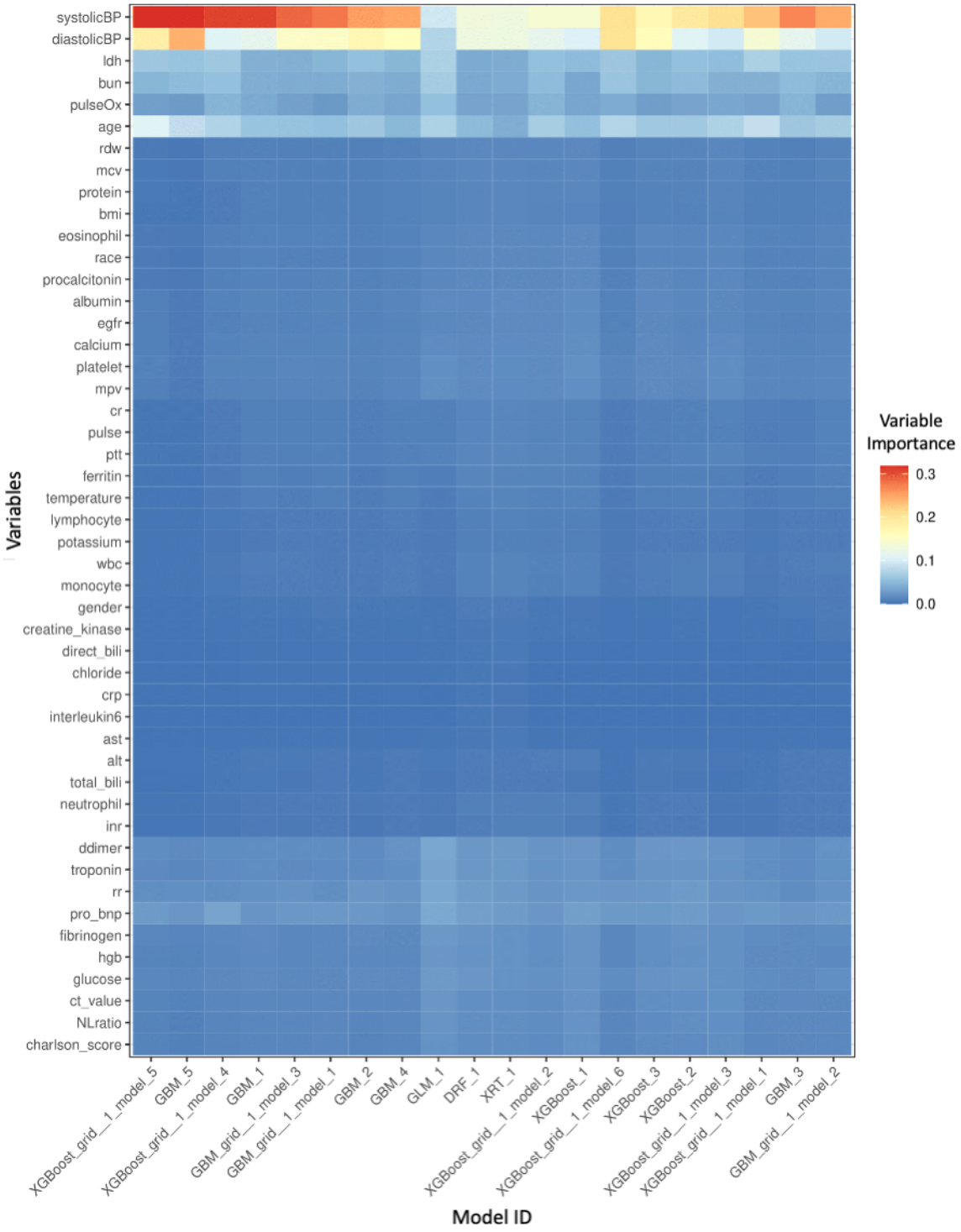


Figure 5. A heatmap that represents the importance of each variable in every machine learning model. DRF: distributed random forest; GBM: gradient boosting machine; GLM: generalized linear model; XGBoost: extreme gradient boosting; XRT: extremely randomized trees.



Selection of the Top 10 Variables: Dimensionality Reduction

The SHAP plots for the GBM and XGBoost models (ie, the two highest-performing models) showed that systolic and diastolic blood pressure, age, LDH level, pulse oximetry level, respiratory rate, BUN level, and troponin level were top 10 variables in both models. BUN and troponin levels are indicators of renal and cardiac function, respectively. With regard to our marker for coagulation, we chose D-dimer level over fibrinogen level because it ranked higher and had more data points. Furthermore, we used Charlson comorbidity scores to represent comorbidities. Glucose level was also a highly ranked variable. This was likely due to the increased risk of mortality in patients with diabetes. However, since the Charlson comorbidity score also accounts for diabetes, we believed that the Charlson comorbidity score

was a more comprehensive predictive variable than glucose level. To confirm whether our choice to use Charlson comorbidity score over glucose level was justified, we trained an autoML model on Charlson comorbidity score data (ie, without glucose level data). We then compared it to a model that was trained on glucose level data (ie, without Charlson comorbidity score data). The former model (AUPRC=0.79) performed better than the latter model (AUPRC=0.78), thereby validating our choice.

According to our variable selection process, the top predictive variables were systolic and diastolic blood pressure, age, LDH level, pulse oximetry level, respiratory rate, BUN level, troponin level, D-dimer level, and Charlson comorbidity score. We believed that these variables provided a good representation of biological processes that are affected by SARS-CoV-2 infection. In addition, these variables are easy to obtain in clinical settings. They also reduce the incidence of missing values.

MODEL-10 Generation and Performance

We used the top 10 influential variables to generate 20 more machine learning models and rank them in order of AUPRC (Table 3). The best performing model was the stacked ensemble of each machine learning algorithm family (AUPRC=0.791). This was MODEL-10. The best performing independent model was the XGBoost model, which had an AUPRC of 0.790.

Table 3. Output of the automated machine learning models that used 10 variables. Model ranks are ordered according to AUPRCs[a].

Rank	Model ID	AUPRC	Area under the curve
1	StackedEnsemble_BestOfFamily_AutoML_20201219_142406	0.791	0.903
2	XGBoost_grid__1_AutoML_20201219_142406_model_6	0.790	0.894
3	StackedEnsemble_AllModels_AutoML_20201219_142406	0.790	0.903
4	GBM_3_AutoML_20201219_142406	0.782	0.898
5	GBM_5_AutoML_20201219_142406	0.782	0.897
6	DRF_1_AutoML_20201219_142406	0.780	0.899
7	XGBoost_grid__1_AutoML_20201219_142406_model_5	0.777	0.893
8	GBM_2_AutoML_20201219_142406	0.777	0.904
9	XGBoost_grid__1_AutoML_20201219_142406_model_1	0.777	0.896
10	XRT_1_AutoML_20201219_142406	0.776	0.900
11	GBM_grid__1_AutoML_20201219_142406_model_1	0.775	0.899
12	XGBoost_grid__1_AutoML_20201219_142406_model_4	0.775	0.894
13	XGBoost_3_AutoML_20201219_142406	0.772	0.891
14	GBM_grid__1_AutoML_20201219_142406_model_2	0.770	0.896
15	GBM_grid__1_AutoML_20201219_142406_model_3	0.770	0.900
16	GBM_1_AutoML_20201219_142406	0.769	0.895
17	XGBoost_2_AutoML_20201219_142406	0.766	0.890
18	GBM_4_AutoML_20201219_142406	0.766	0.897
19	XGBoost_1_AutoML_20201219_142406	0.762	0.886
20	XGBoost_grid__1_AutoML_20201219_142406_model_3	0.761	0.885
21	XGBoost_grid__1_AutoML_20201219_142406_model_2	0.754	0.889
22	GLM_1_AutoML_20201219_142406	0.733	0.860

[a]AUPRC: area under the precision-recall curve.

Performance of MODEL-48 and MODEL-10 as Binary Classifiers

The maximum F2 score of MODEL-48 was 0.793, and the probability threshold was 0.110. The binary classifier for this threshold had a sensitivity, specificity, PPV, and NPV of 0.919, 0.735, 0.513, and 0.968, respectively (Figure 6). The maximum F2 score of MODEL-10 was 0.779, and the probability threshold was 0.202. The binary classifier for this threshold had a sensitivity, specificity, PPV, and NPV of 0.838, 0.836, 0.609, and 0.944, respectively (Figure 7).

Figure 6. Binary classifier of MODEL-48. The model had an optimized F2-score threshold. This classifier had a sensitivity of 0.92, a specificity of 0.74, a positive predictive value of 0.51, and a negative predictive value of 0.97.

Predicted outcome of MODEL-48	True Outcome	
	Dead, n	Alive, n
Dead, n	181	172
Alive, n	16	476

Figure 7. Binary classifier of MODEL-10. The model had an optimized F2-score threshold. This classifier had a sensitivity of 0.84, a specificity of 0.84, a positive predictive value of 0.61, and a negative predictive value of 0.94.

Predicted outcome of MODEL-10	True Outcome	
	Dead, n	Alive, n
Dead, n	165	106
Alive, n	32	542

Discussion

Principal Findings

We were able to use autoML and clinical values (ie, those that were collected early during a patient's admission to a hospital) to successfully generate multiple machine learning models, assess their performance, and select the highest-performing models for predicting patients' chances of surviving a SARS-CoV-2 infection. In addition, our study demonstrates that machine learning models that only use 10 clinical variables can predict survival. These models also had high sensitivity values, specificity values, and NPVs. Therefore, autoML is an efficient, informative, and easily reproducible method. The clinical implementation of autoML-based models may require further investigation. However, we demonstrate that autoML is a comprehensive approach for building machine learning–based clinical decision support tools.

Our results show that the best models were GBM and XGBoost models. They both had high performance, as determined by their AUPRCs and AUCs. The RF model, DL model, and GLM performed substantially worse compared to the GBM and XGBoost models. The DL model may have performed better if we had a larger data set, but our DL model required much longer training times than the other models. Tree-based machine learning algorithms (eg, GBM, XGBoost, and RF) are more efficient, and possibly more effective, than neural network algorithms in terms of analyzing tabular data. We used the AUPRC as our metric of model utility because it accounts for the two critical clinical performance metrics that were of specific interest to us—the positive predictive value and sensitivity. We wanted to identify patients who were likely to die so that we could take action and treat as many patients as possible. Alternatively, the AUC accounts for model sensitivity and specificity and ignores the effects of mortality prevalence on model performance. The prevalence of mortality sets the context in which the model must perform; without this information, the model is irrelevant.

Machine learning models can be used to enhance electronic medical record systems and calculate the values of variables that are collected from patients. Based on the performance of MODEL-10, our dimensionality reduction process was successful; 10 variables were enough to generate high-performing models. This shows that not all parameters are necessary for performing calculations and making predictions. Clinicians and hospitals should begin the patient assessment process by prioritizing the ordering of medical tests (ie, tests for the 10 variables). Dimensionality reduction not only reduced the number of variables we needed to consider, but also minimized the number of missing values in the data set and reduced the risk of imputation bias. This may be the reason why the performance of MODEL-10 was similar to that of MODEL-48. These 10 variables may also help researchers with conducting studies on unique cohorts and reproducing our results.

The purpose of autoML is not limited to predicting the survival of patients with COVID-19. AutoML can be used to generate models that are based on other types of clinical data and predict other outcomes (eg, models for predicting which patients require a ventilator). We hope that our study helps other researchers with applying our autoML approach, accelerating the implementation of artificial intelligence models into medical systems, and delivering better medical care.

Clinical Insights From the Black Box

The trade-off between predictive power and interpretability is a common issue when working with black-box models, especially in medical environments where results have to be explained to medical providers and patients. Interpretability is crucial for questioning, understanding, and trusting artificial intelligence and machine learning systems. According to our variable importance heatmap (Figure 5), many models

determined that age, pulse oximetry level, and systolic and diastolic blood pressure were important variables for predicting the outcome. Biomarkers such as BUN level, LDH level, estimated glomerular filtration rate, and probrain natriuretic peptide level were also influential variables. These findings show that our model results are in line with the clinical findings of other researchers [23-27].

Our SHAP and PD plots provided insight into the black box. The SHAP plots allowed us to determine the importance of variables and provided information on how the variables influenced models' predictions. Alternatively, PD plots provide numerical information on a variable's effects. For example, the SHAP plots for the GBM and XGBoost models showed that high glucose levels were associated with an increased probability of mortality (ie, high SHAP value). Additionally, the PD plots showed that increases in glucose level were proportional to increases in patients' chances of death. However, this was only true for glucose levels of <300 mg/dL. These results may support the idea that people with diabetes are at an increased risk of mortality. Therefore, SHAP and PD plots can be used to confirm clinical findings and show clinical thresholds.

Other variables that were less influential are also worthy of examination. For example, the SHAP plots (Figure 3) for the GBM model showed that low albumin levels were weakly associated with an increased chance of death. Such findings may provide insight into the disease mechanism of COVID-19. In addition, models that only use 10 variables can be implemented by institutions that might not be able to collect all 48 variables that were tested in this study. If high-performing models can be generated with only 10 variables, clinicians and hospitals can focus on collecting these variables when conducting patient assessments. Such models also minimize the problem of data sparsity and the risk of imputing missing data points. In addition, the use of such models will help clinicians with manually entering values (eg, inputting values on a mobile device).

Limitations and Future Work

We recognize that there were limitations to our study. Our cohort was limited to patients with severe conditions that required them to be admitted to a hospital. Therefore, our findings may not be generalizable to all patients with COVID-19. For example, we were surprised to learn that our machine learning models did not identify race as an important predictor of death, given the fact that at a population level, the relative risk of mortality among Black patients is higher than that of White patients [28]. However, one population study analyzed the relative risk of mortality among Black patients and White patients with COVID-19 who were admitted to our institution (ie, Montefiore Medical Center). This study found that there were no considerable differences in mortality rates between the two groups once patients were admitted to our hospital [28]. During a pandemic, hospital beds are scarce. Therefore, only patients who exhibit severe symptoms are allowed to be admitted to emergency rooms. Fortunately, our colleagues from our institution conducted a conventional logistic regression analysis, which resulted in the same finding; hospitalized White and Black patients with illnesses of equal severity and relevant predictor's for disease progression at admission had similar mortality rates [28]. The results from our machine learning models are in line with those of the logistic regression analysis.

In our study, systolic and diastolic blood pressure were the most important variables. However, these variables may simply indicate that patients with severe illnesses and hypotension are at an imminent risk of death. Temporal features were not considered in our analysis. For example, we did not determine whether hypotension at admission was an important variable for patients who survive during the first 24 hours of admission. Further, we did not determine whether a variable's importance diminishes in populations that survive after the first 48 or 72 hours of admission. In a future study, we would like to test whether our models are robust enough to predict death during different times of admission. For example, we would test our models' performance for predicting the death of patients within the first week of admission and the fourth week after admission.

The handling of missing values is a challenging problem in machine learning model development. Fibrinogen, procalcitonin, and cycle threshold values were missing for many people in our cohort (Table S1 in Multimedia Appendix 1). We understand that missing values are not indicative of a variable's clinical importance. For example, changes in practice patterns alter the meaning of missing data. After the first few weeks of the COVID-19 pandemic, our institution implemented a best practice protocol that required clinicians to measure patients' D-dimer levels upon admission and recommend anticoagulation treatment to patients with elevated D-dimer levels. Clearly, the presence of D-dimers during the early period of the pandemic had a different meaning from that of the presence of D-dimers during the later period of the pandemic. Similarly, missing D-dimer level values were considered random events during the later period of the pandemic and purposeful events during the early period of the pandemic [29]. Therefore, missing data from the early period of the pandemic have a different meaning compared to that of missing data from the later period. However, our imputation methods did not account for temporal changes in the meaning of missing data. This is an important challenge that must be considered in future machine learning software development studies.

With regard to the generalization of our model, future studies need to be conducted to assess whether our models are useful at other institutions during the second wave of COVID-19. As patient demographics can differ from institution to institution, hospitals may need to customize their models in accordance with their patient populations. Models should also be designed to integrate new data and adjust to the ever-changing environment. We are continually working on reinforcement learning methods for updating our model in real time.

Conclusion

We used autoML to generate high-performing machine learning models that predicted the mortality of patients with COVID-19. We also identified important variables that were strongly associated with patients' survival. Our study provides proof of concept that autoML is an efficient, effective, and informative method for training machine learning models and gaining insight into disease processes. AutoML models may help clinicians

with triaging patients during the COVID-19 pandemic. Our COVID-19 mortality calculator, which is based on this study, is freely available online as a web-based computer application [5].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of all variables that were collected from the entire cohort and used in the training and testing data sets.
[PDF File (Adobe PDF File), 181 KB - jmir_v23i2e23458_app1.pdf]

Multimedia Appendix 2
Partial dependence plots of the most important variables.
[PDF File (Adobe PDF File), 317 KB - jmir_v23i2e23458_app2.pdf]

References

1. COVID data tracker. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html [accessed 2021-02-12]
2. Cheng FY, Joshi H, Tandon P, Freeman R, Reich DL, Mazumdar M, et al. Using machine learning to predict ICU transfer in hospitalized COVID-19 patients. J Clin Med 2020 Jun 01;9(6):1668 [FREE Full text] [doi: 10.3390/jcm9061668] [Medline: 32492874]
3. Vaid A, Somani S, Russak AJ, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine learning to predict mortality and critical events in a cohort of patients with COVID-19 in New York City: Model development and validation. J Med Internet Res 2020 Nov 06;22(11):e24018 [FREE Full text] [doi: 10.2196/24018] [Medline: 33027032]
4. Yan L, Zhang HT, Goncalves J, Xiao Y, Wang M, Guo Y, et al. An interpretable mortality prediction model for COVID-19 patients. Nat Mach Intell 2020 May 14;2:283-288 [FREE Full text] [doi: 10.1038/s42256-020-0180-7]
5. Montefiore Medical Center. Tsubomi Technology. URL: https://www.tsubomi.tech/ [accessed 2021-02-23]
6. Castro VM, McCoy TH, Perlis RH. Laboratory findings associated with severe illness and mortality among hospitalized individuals with coronavirus disease 2019 in eastern Massachusetts. JAMA Netw Open 2020 Oct 01;3(10):e2023934 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.23934] [Medline: 33125498]
7. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, China Medical Treatment Expert Group for COVID-19. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. JAMA Intern Med 2020 Aug 01;180(8):1081-1089 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2033] [Medline: 32396163]
8. Kermali M, Khalsa RK, Pillai K, Ismail Z, Harky A. The role of biomarkers in diagnosis of COVID-19 - A systematic review. Life Sci 2020 Aug 01;254:117788 [FREE Full text] [doi: 10.1016/j.lfs.2020.117788] [Medline: 32475810]
9. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
10. Ruan Q, Yang K, Wang W, Jiang L, Song J. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med 2020 May;46(5):846-848 [FREE Full text] [doi: 10.1007/s00134-020-05991-x] [Medline: 32125452]
11. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. Int J Infect Dis 2020 May;94:91-95 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.017] [Medline: 32173574]
12. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clin Chem Lab Med 2020 Jun 25;58(7):1021-1028. [doi: 10.1515/cclm-2020-0369] [Medline: 32286245]
13. Bellin E. Riddles in Accountable Healthcare: A Primer to Develop Analytic Intuition for Medical Homes and Population Health. North Charleston, South Carolina: Create Space; Mar 07, 2015.
14. Bellin E. How to Ask and Answer your Research Question using Electronic Medical Record Data. North Charleston, South Carolina: Create Space; 2017.
15. Bellin E. Missing Management: Health-Care Analytic Discovery in a Learning Health System. Seattle, Washington: Kindle Direct Publishing; Sep 05, 2019.
16. Bellin E, Fletcher DD, Geberer N, Islam S, Srivastava N. Democratizing information creation from health care data for quality improvement, research, and education-the Montefiore Medical Center Experience. Acad Med 2010 Aug;85(8):1362-1368. [doi: 10.1097/ACM.0b013e3181df0f3b] [Medline: 20453810]
17. h2oai / h2o-3. GitHub. URL: https://github.com/h2oai/h2o-3 [accessed 2021-02-12]

18. AutoML: Automatic machine learning. H2O.ai. URL: http://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html [accessed 2021-02-12]
19. The R Project for statistical computing. R Foundation. URL: https://www.R-project.org/ [accessed 2021-02-12]
20. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020 Jan;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
21. Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat 2001;29(5):1189-1232. [doi: 10.1214/aos/1013203451]
22. Molnar C. Interpretable machine learning: A guide for making black box models explainable. GitHub. 2021 Feb 08. URL: https://christophm.github.io/interpretable-ml-book/ [accessed 2021-02-12]
23. Pei G, Zhang Z, Peng J, Liu L, Zhang C, Yu C, et al. Renal involvement and early prognosis in patients with COVID-19 pneumonia. J Am Soc Nephrol 2020 Jun;31(6):1157-1165. [doi: 10.1681/ASN.2020030276] [Medline: 32345702]
24. Deng Q, Hu B, Zhang Y, Wang H, Zhou X, Hu W, et al. Suspected myocardial injury in patients with COVID-19: Evidence from front-line clinical observation in Wuhan, China. Int J Cardiol 2020 Jul 15;311:116-121 [FREE Full text] [doi: 10.1016/j.ijcard.2020.03.087] [Medline: 32291207]
25. Thachil J, Tang N, Gando S, Falanga A, Cattaneo M, Levi M, et al. ISTH interim guidance on recognition and management of coagulopathy in COVID-19. J Thromb Haemost 2020 May;18(5):1023-1026. [doi: 10.1111/jth.14810] [Medline: 32338827]
26. Zhang Y, Xiao M, Zhang S, Xia P, Cao W, Jiang W, et al. Coagulopathy and antiphospholipid antibodies in patients with covid-19. N Engl J Med 2020 Apr 23;382(17):e38 [FREE Full text] [doi: 10.1056/NEJMc2007575] [Medline: 32268022]
27. Wang S, Ma P, Zhang S, Song S, Wang Z, Ma Y, et al. Fasting blood glucose at admission is an independent predictor for 28-day mortality in patients with COVID-19 without previous diagnosis of diabetes: a multi-centre retrospective study. Diabetologia 2020 Oct;63(10):2102-2111 [FREE Full text] [doi: 10.1007/s00125-020-05209-1] [Medline: 32647915]
28. Golestaneh L, Neugarten J, Fisher M, Billett HH, Gil MR, Johns T, et al. The association of race and COVID-19 mortality. EClinicalMedicine 2020 Aug;25:100455 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100455] [Medline: 32838233]
29. Billett HH, Reyes-Gil M, Szymanski J, Ikemura K, Stahl LR, Lo Y, et al. Anticoagulation in COVID-19: Effect of enoxaparin, heparin, and apixaban on mortality. Thromb Haemost 2020 Dec;120(12):1691-1699. [doi: 10.1055/s-0040-1720978] [Medline: 33186991]

Abbreviations

AUC: area under the curve
AUPRC: area under the precision-recall curve
autoML: automated machine learning
BUN: blood urea nitrogen
CLG: Clinical Looking Glass
DL: deep learning
GLM: general linear model
GBM: gradient boosting machine
LDH: lactate dehydrogenase
NPV: negative predictive value
PD: partial dependence
PPV: positive predictive value
RF: random forest
SHAP: Shapley additive explanation
XGBoost: extreme gradient boosting

Edited by C Basch; submitted 17.08.20; peer-reviewed by S Nagavally, A James, A Kardos; comments to author 02.12.20; revised version received 23.12.20; accepted 03.02.21; published 26.02.21.

Please cite as:
Ikemura K, Bellin E, Yagi Y, Billett H, Saada M, Simone K, Stahl L, Szymanski J, Goldstein DY, Reyes Gil M
Using Automated Machine Learning to Predict the Mortality of Patients With COVID-19: Prediction Model Development Study
J Med Internet Res 2021;23(2):e23458
URL: https://www.jmir.org/2021/2/e23458
doi:10.2196/23458
PMID:33539308

Original Paper

Influence of Health Beliefs on Adherence to COVID-19 Preventative Practices: International, Social Media–Based Survey Study

Julianna C Hsing[1,2], BA; Jasmin Ma[2], BS; Alejandra Barrero-Castillero[3,4,5], MD, MPH; Shilpa G Jani[2], MPH; Uma Palam Pulendran[2], MPH; Bea-Jane Lin[2], PhD; Monika Thomas-Uribe[6], MD, MPH; C Jason Wang[2,7,8], MD, PhD

[1]Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, United States

[2]Center for Policy, Outcomes, and Prevention, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, United States

[3]Division of Neonatology, Beth Israel Deaconess Medical Center, Boston, MA, United States

[4]Division of Newborn Medicine, Boston Children's Hospital, Boston, MA, United States

[5]Department of Pediatrics, Harvard Medical School, Boston, MA, United States

[6]Department of Pediatrics, University of California San Francisco - Fresno, Fresno, CA, United States

[7]Center for Health Policy, Freeman-Spogli Institute for International Studies, Stanford University, Stanford, CA, United States

[8]Center for Primary Care Outcomes Research, Stanford University School of Medicine, Stanford, CA, United States

Corresponding Author:
C Jason Wang, MD, PhD
Center for Policy, Outcomes, and Prevention
Department of Pediatrics
Stanford University School of Medicine
117 Encina Commons, CHP/PCOR
Stanford, CA, 94305
United States
Phone: 1 (650) 736 0403
Email: cjwang1@stanford.edu

Abstract

Background: Health behavior is influenced by culture and social context. However, there are limited data evaluating the scope of these influences on COVID-19 response.

Objective: This study aimed to compare handwashing and social distancing practices in different countries and evaluate practice predictors using the health belief model (HBM).

Methods: From April 11 to May 1, 2020, we conducted an online, cross-sectional survey disseminated internationally via social media. Participants were adults aged 18 years or older from four different countries: the United States, Mexico, Hong Kong (China), and Taiwan. Primary outcomes were self-reported handwashing and social distancing practices during COVID-19. Predictors included constructs of the HBM: perceived susceptibility, perceived severity, perceived benefits, perceived barriers, self-efficacy, and cues to action. Associations of these constructs with behavioral outcomes were assessed by multivariable logistic regression.

Results: We analyzed a total of 71,851 participants, with 3070 from the United States, 3946 from Mexico, 1201 from Hong Kong (China), and 63,634 from Taiwan. Of these countries, respondents from the United States adhered to the most social distancing practices (χ^2_3=2169.7, *P*<.001), while respondents from Taiwan performed the most handwashing (χ^2_3=309.8, *P*<.001). Multivariable logistic regression analyses indicated that self-efficacy was a positive predictor for handwashing (odds ratio [OR]$_{\text{United States}}$ 1.58, 95% CI 1.21-2.07; OR$_{\text{Mexico}}$ 1.5, 95% CI 1.21-1.96; OR$_{\text{Hong Kong}}$ 2.48, 95% CI 1.80-3.44; OR$_{\text{Taiwan}}$ 2.30, 95% CI 2.21-2.39) and social distancing practices (OR$_{\text{United States}}$ 1.77, 95% CI 1.24-2.49; OR$_{\text{Mexico}}$ 1.77, 95% CI 1.40-2.25; OR$_{\text{Hong Kong}}$ 3.25, 95% CI 2.32-4.62; OR$_{\text{Taiwan}}$ 2.58, 95% CI 2.47-2.68) in all countries. Handwashing was positively associated with perceived susceptibility in Mexico, Hong Kong, and Taiwan, while social distancing was positively associated with perceived severity in the United States, Mexico, and Taiwan.

Conclusions: Social media recruitment strategies can be used to reach a large audience during a pandemic. Self-efficacy was the strongest predictor for handwashing and social distancing. Policies that address relevant health beliefs can facilitate adoption of necessary actions for preventing COVID-19. Our findings may be explained by the timing of government policies, the number of cases reported in each country, individual beliefs, and cultural context.

(J Med Internet Res 2021;23(2):e23720) doi:10.2196/23720

KEYWORDS

COVID-19 pandemic; health belief model; behavior change; preventative health behaviors; handwashing; social distancing; international; online survey; social media; cross-sectional study

Introduction

The severity and rapid transmission of COVID-19 has forced most regions to implement community mitigation strategies. These strategies have ranged from government guidelines on personal protective measures and social distancing to strict lockdown orders that closed schools and businesses [1]. Nationwide school closures in 194 countries in early April 2020 demonstrated the extent of these interventions [2]. These measures have reduced transmission or delayed the peak of infection of past pandemics to varying degrees, which were estimated to have prevented at least 60 million COVID-19 cases [3,4].

Although these interventions reduce the stress on health care systems, they also incur high economic and societal costs, making adherence more difficult for those under financial strain [3,5]. Recent studies have begun to assess adherence to COVID-19 guidelines, evaluating demographic characteristics and the impact of guideline duration [6-8]. Some have suggested that concepts from social and behavioral sciences can provide insight into adherence to guidelines, but current data evaluating these hypotheses in multiple countries and in the context of COVID-19 using relevant behavior change theories, such as the health belief model (HBM), are limited [9]. Given the rapid spread of COVID-19 and the scale of guidelines worldwide, a cross-cultural assessment of preventative health behaviors is essential to identifying which approaches improve adherence. This study aims to compare handwashing and social distancing behaviors across four different countries using the HBM.

Methods

Participant Recruitment

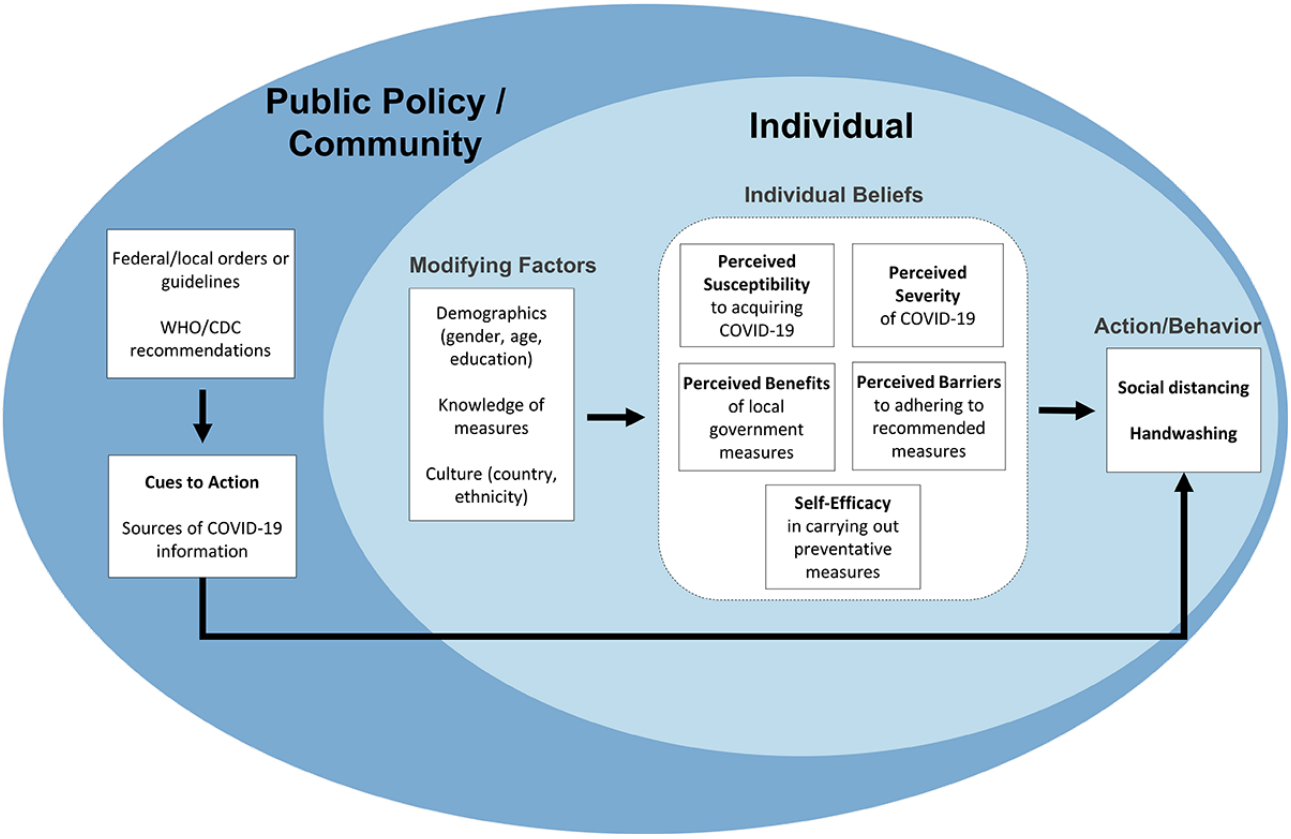
From April 11 to May 1, 2020, we conducted a confidential, cross-sectional, international open survey through the following social media platforms: Facebook, Instagram, Line, and Twitter. The survey was announced and advertised through Stanford Health Policy's social media accounts. Facebook boosted posts were used to target social media users who were 18 years of age or older. We focused our analysis on countries and regions with at least 1000 survey responses: the United States, Mexico, Hong Kong (China), and Taiwan. Facebook is the most popular social media platform among adults in all four countries, whereas Instagram, Twitter, and Line have relatively high penetration in specific groups and countries [10-13]. Though the limitations of using convenience sampling and social media are well-known, this method is cost-effective, time-efficient, and most feasible for reaching a large international audience in a fast-spreading pandemic [14]. The alternative of administering telephone surveys is associated with extremely low response rates (6% in 2018) and limitations on item complexity and survey length [15,16].

The survey was developed on Qualtrics (Qualtrics Inc), an online survey distribution tool, and administered in English, Spanish, and Mandarin. Translations were provided by native speakers fluent in the respective languages, who tested the survey before it was fielded. Prior to survey completion, participants were provided with information about the study and were asked to acknowledge consent to the study. All items were optional except for country of residence. Through Qualtrics, cookies were used to assign a unique user identifier to each client computer to prevent participants from completing the survey more than once. Only completed surveys were analyzed. Given that no incentives were offered to participants and that the survey was voluntary, we did not assess whether surveys were completed in an atypical amount of time. The study was reviewed and approved by Stanford University's Institutional Review Board.

Conceptual Model and Survey Items

We used the HBM, a widely used framework for explaining health behaviors and guiding related interventions, to create survey items to assess health beliefs among respondents in the four countries [17]. The HBM accounts for both individual- and community-level factors of health motivation, making it an ideal option for addressing health behavior problems that evoke health concerns during the COVID-19 pandemic. Figure 1 shows the key constructs in the HBM that determine behavior, including individual beliefs (ie, perceived susceptibility, severity, benefits, and barriers as well as self-efficacy), which may be influenced by sociodemographic factors or knowledge, and cues to action, which may be influenced by public policy.

Figure 1. Conceptual framework of the study adapted from the health belief model to assess individual health beliefs, modifying factors, and the effects of public policy on social distancing and handwashing behaviors during the COVID-19 pandemic. CDC: Centers for Disease Control and Prevention; WHO: World Health Organization.



Survey items (see Multimedia Appendix 1) were developed based on prior expertise, survey knowledge, and group discussion. We asked participants the following questions for each HBM construct:

1. Perceived susceptibility. What do you think your risk is of getting infected with COVID-19?
2. Perceived severity. How afraid are you of the COVID-19 pandemic?
3. Perceived benefits. How do you feel about the government measures of COVID-19 in your area?
4. Perceived barriers. Have any barriers prevented you from adhering to measures in your area?
5. Self-efficacy. How confident are you that you are able to and willing to carry out these measures?
6. Cues to action. What are your sources of information regarding COVID-19?

Perceived susceptibility, perceived severity, and self-efficacy items were assessed using a 5-point Likert scale. During analysis, response scales were eventually collapsed into three categories, such as *not likely/slightly likely*, *moderately likely*, and *likely/very likely*. Perceived benefit items were assessed using three categories: *unnecessarily restrictive/moderately restrictive*, *essential/appropriate*, and *not enough*. Perceived barriers and cues to action were both assessed on a binary *yes/no* scale.

To account for modifying factors that influenced individual beliefs, we assessed for age (ie, 18-24, 25-34, 35-44, 45-59, and 60+ years), gender (ie, male, female, and other), highest educational attainment (ie, high school or less and college and above), country of residence (ie, United States, Mexico, Hong Kong, and Taiwan), race or ethnicity (eg, Asian, Hispanic/Latino, and White or European), change in income due to COVID-19 (ie, yes or no), and awareness of government measures or guidelines (ie, some/not aware and most/all).

Handwashing behaviors were assessed by asking respondents whether they washed their hands or used hand sanitizer in the following seven situations: (1) after coming home from being outside; (2) after grocery shopping; (3) after interacting with nonhousehold members; (4) while being in public; (5) before or after using their vehicle; (6) after blowing their nose, coughing, or sneezing into their hand; and (7) before eating. Responses for all situations were summed up to a score of 7. Social distancing behaviors were evaluated by assessing whether respondents did the following: (1) avoided nonessential gatherings, (2) kept at least the recommended distance from nonhousehold members (eg, 6 feet, 1.5 meters, 2 meters, etc), or (3) avoided close contact with individuals at higher risk for severe illness from COVID-19. Responses were summed up to a score of 3. Total adherence to either handwashing or social distancing responses was assessed by a binary variable, with individuals performing all of the practices as one group (yes = 1) and those who performed fewer than all practices as the other group (no = 0); this was done for each of the two behaviors.

Statistical Analysis

We conducted poststratification weighting for each country by age and gender—and race or ethnicity for the United

States—using each country's most recent census data [18-21]. Weights were calculated by dividing each stratified proportion of the country's population by each stratified proportion of the study's country sample, followed by renormalizing for each country to ensure that weighted sample size equaled the unweighted sample size [22]. Weighted frequencies and percentages were calculated for categorical variables and compared using chi-square tests. To assess which country performed more handwashing and social distancing practices, countries were analyzed together in multivariate analyses, with each country coded as a key independent dichotomous variable, adjusting for gender, age, education, and reduced income.

Countries were also analyzed separately with multivariable logistic regressions to examine the association of HBM constructs with two main outcomes: handwashing and social distancing practices. HBM covariates included perceived susceptibility, severity, benefits, and barriers; self-efficacy; and cues to action. All models were adjusted for gender, age, education, and reduced income. To ensure our handwashing variable appropriately captured COVID-19-related handwashing behaviors, we also ran a sensitivity analysis that assessed the association between handwashing time (ie, >20 seconds vs ≤20 seconds) and HBM constructs, because this handwashing duration was a specific COVID-19 recommendation in all four countries [23-26]. For all models, odds ratios (ORs) and 95% CIs were calculated. All statistical analyses were performed using R statistical software, version 3.6.3 (The R Foundation), and *P* values were 2-sided with an α of .05.

Results

Participant Characteristics

A total of 71,851 individuals were included in our analysis: 3070 from the United States (4.3%), 3946 from Mexico (5.5%), 1201 from Hong Kong (1.7%), and 63,634 from Taiwan (88.6%). Of these, 71,728 (99.8%) completed at least 80% of the survey (14 of 17 questions). Missing data for each item were less than 5% and, thus, were not imputed. After weighting, the gender and age distributions were representative of each country according to their most recent census data (see Multimedia Appendices 2 and 3). Overall, Mexico had a younger population compared to other countries, and most respondents in all four countries had a college degree or higher. A total of 2099 out of 3931 respondents (53.4%) from Mexico, 337 out of 1189 (28.3%) from Hong Kong, 779 out of 3062 (25.4%) from the United States, and 10,725 out of 63,399 (16.9%) from Taiwan reported reduced income due to COVID-19. A total of 68,614 out of 71,633 respondents (95.8%) in all countries were aware of government measures and/or guidelines (see Table 1).

Table 1. Weighted demographic characteristics of survey respondents by country.

Characteristic	Value (N=71,851), n (%)[a]				*P* value[b]
	United States (n=3070)	Mexico (n=3946)	Hong Kong (n=1201)	Taiwan (n=63,634)	
Age group (years)					<.001
18-24	110 (3.6)	507 (12.9)	83 (7.0)	4969 (7.8)	—[c]
25-34	519 (17.0)	953 (24.2)	198 (16.6)	10,509 (16.6)	—
35-44	451 (14.7)	820 (20.8)	226 (18.9)	12,865 (20.3)	—
45-59	963 (31.4)	977 (24.8)	310 (25.9)	17,836 (28.1)	—
60+	1019 (33.3)	684 (17.3)	377 (31.6)	17,294 (27.2)	—
Gender					<.001
Female	1683 (55.0)	2031 (51.6)	602 (50.4)	31,407 (49.6)	—
Male	1351 (44.2)	1867 (47.4)	562 (47.1)	30,034 (47.4)	—
Other[d]	25 (0.8)	40 (1.0)	30 (2.5)	1894 (3.0)	—
Race or ethnicity					<.001
Asian	158 (5.2)	15 (0.4)	1180 (98.9)	62,924 (99.3)	—
Hispanic/Latino or other[e]	1520 (49.7)	3319 (84.7)	11 (0.9)	228 (0.4)	—
White or European	1379 (45.1)	586 (15.0)	2 (0.2)	207 (0.3)	—
Education					<.001
Below college	286 (9.4)	726 (18.4)	263 (22.2)	7889 (12.4)	—
College and above	2768 (90.6)	3215 (81.6)	923 (77.8)	55,547 (87.6)	—
Reduced income since COVID-19					<.001
No	2283 (74.6)	1832 (46.6)	852 (71.7)	52,674 (83.1)	—
Yes	779 (25.4)	2099 (53.4)	337 (28.3)	10,725 (16.9)	—
Awareness of governmental measures and/or guidelines					<.001
Some/not aware	20 (0.7)	123 (3.1)	26 (2.2)	2850 (4.5)	—
All/most	3034 (99.3)	3818 (96.9)	1167 (97.8)	60,595 (95.5)	—

[a]Weighted values were calculated by dividing the actual proportion of the country's population by the proportion from the study's sample, then renormalized for each country to ensure weighted and unweighted sample sizes were equal. Due to rounding and missing data (<5% for each item), the sum of frequencies and percentages for the sample weighted columns may not equal the country's total sample size.

[b]*P* values were calculated using 2-sided chi-square tests.

[c]Not available.

[d]Responses of *other* gender include individuals who chose nonbinary/third gender, prefer not to say, or other (<3% of total responses).

[e]Responses of *other* race or ethnicity include individuals who are African, Black, African American, American Indian or Alaskan Native, Middle Eastern, Native Hawaiian or other Pacific Islander, or other. Categories were collapsed due to low numbers (<2% of total responses).

Handwashing and Social Distancing Behaviors

Bivariate chi-square analyses showed that respondents from Taiwan practiced the most handwashing behaviors (χ^2_3=309.8, *P*<.001) relative to other countries, while those from the United States practiced the most social distancing (χ^2_3=2169.7, *P*<.001). Of the 71,608 respondents who provided a response to their handwashing practices, 39.6% (1215/3066) from the United States, 48.8% (1927/3938) from Mexico, 45% (538/1195) from Hong Kong, and 54.1% (34,328/63,409) from Taiwan reported handwashing in all seven situations (see Figure 2). Of the 71,851 respondents who provided a response to their social distancing practices, 88.0% (2702/3070) from the United States, 64.3% (2539/3946) from Mexico, 44.7% (537/1201) from Hong Kong, and 48.3% (30,737/63,634) from Taiwan reported performing all three social distancing practices (see Figure 3). We found similar patterns of association in the sensitivity multivariate analysis (see Multimedia Appendix 4). Respondents from the United States (OR 0.50, 95% CI 0.46-0.54), Mexico (OR 0.87, 95% CI 0.81-0.93), and Hong Kong (OR 0.67, 95% CI 0.59-0.75) were less likely to perform handwashing compared to those from Taiwan (reference value). Respondents from the United States (OR 7.73, 95% CI 6.93-8.66) and Mexico (OR 2.17, 95% CI 2.02-2.33) were more likely to practice social distancing compared to those from Taiwan. In contrast, respondents from Hong Kong were less likely to practice social

distancing compared to those from Taiwan, although the association was only slightly significant (OR 0.88, 95% CI 0.87-0.99).

Figure 2. Distribution of handwashing practices by country. Respondents were asked whether they washed their hands or used hand sanitizer in any of the following seven situations: (1) after coming home from being outside; (2) after grocery shopping; (3) after interacting with nonhousehold members; (4) while being in public; (5) before or after using their vehicle; (6) after blowing their nose, coughing, or sneezing into their hand; and (7) before eating.

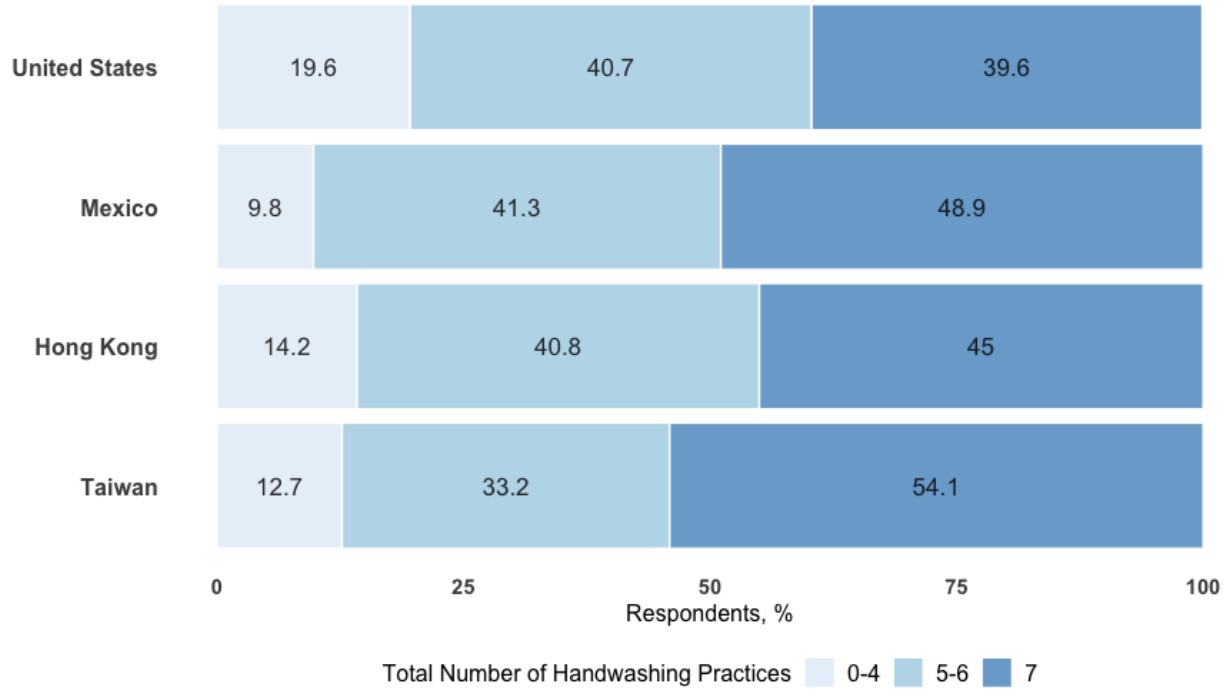
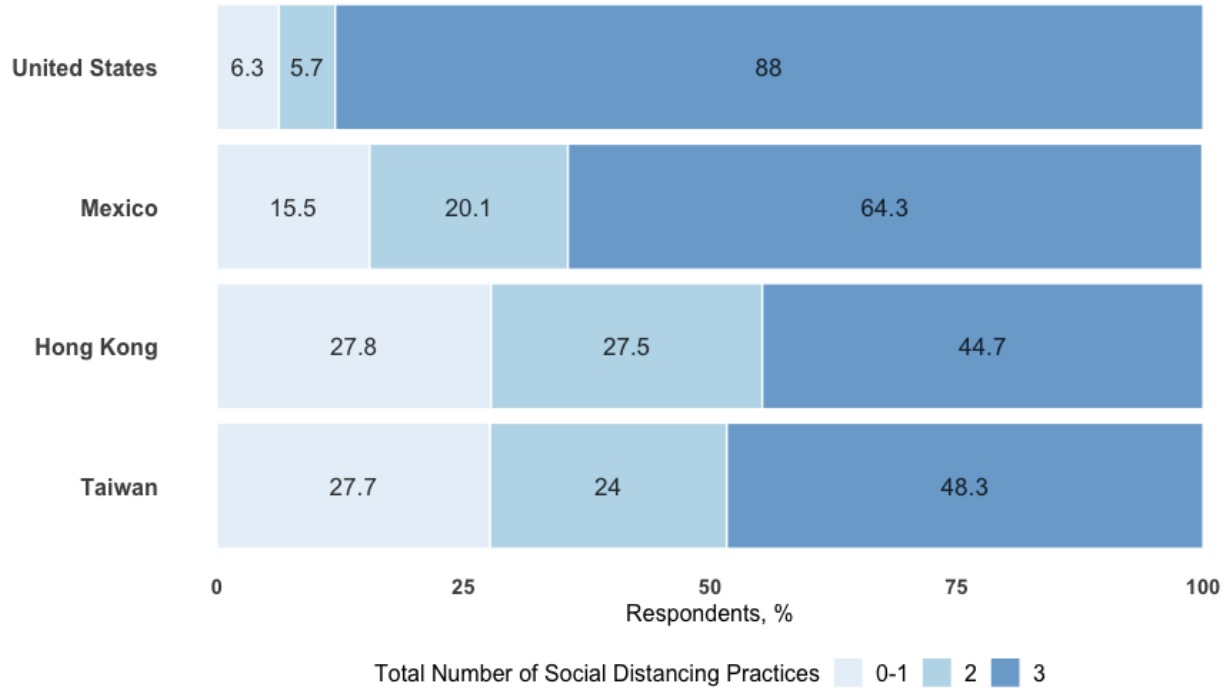


Figure 3. Distribution of social distancing practices by country. Respondents were asked whether they did the following: (1) avoided nonessential gatherings, (2) kept at least the recommended healthy distance from nonhousehold members (eg, 6 feet, 1.5 meters, and 2 meters), or (3) avoided close contact with individuals at higher risk for severe illness from COVID-19.



Health Belief Model Constructs

Table 2 compares the distribution of responses to health belief questions by country, which were assessed using chi-square tests and were all statistically different across each country (P<.001). For perceived susceptibility, the percentage of respondents who felt they were likely (ie, moderately likely to very likely) to be infected with COVID-19 was higher in the United States (1683/3068, 54.9%) and Mexico (2688/3941, 68.2%) compared to those from Hong Kong (557/1192, 46.7%) and Taiwan (19,080/63,425, 30.1%). For perceived severity, a higher percentage of individuals from Taiwan (28,082/63,470, 44.2%) were not afraid (ie, slightly afraid or not afraid) of the COVID-19 pandemic compared to those from the United States (876/3064, 28.6%), Mexico (1154/3940, 29.3%), and Hong Kong (289/1195, 24.2%). For perceived benefits, most individuals in the United States (2062/3016, 68.4%) and Taiwan (53,573/62,625, 85.5%) believed that the government measures in place were appropriate or essential, while most individuals in Mexico (2256/3871, 58.3%) and Hong Kong (745/1172, 63.6%) believed that their measures were not enough. For self-efficacy, a majority of individuals in all countries were confident or very confident in their ability to carry out preventative measures: 88.5% (2712/3066) in the United States, 91.4% (3599/3940) in Mexico, 79.3% (946/1193) in Hong Kong, and 75.8% (48,144/63,481) in Taiwan.

Regarding perceived barriers to social distancing, Mexico (2547/3946, 64.5%) had the highest proportion of individuals who perceived difficulty in obtaining face masks, followed by 60.5% (1856/3070) of individuals in the United States, 52.7% (633/1201) in Hong Kong, and 12.2% (7736/63,634) in Taiwan. Having an essential job (eg, grocery store worker), as perceived by the individual or determined by local governments, was a common perceived barrier in all countries. Other common barriers included family obligations in Mexico as well as transportation needs in Hong Kong and Taiwan. For handwashing barriers, respondents from the United States (1536/3070, 50.0%) and Mexico (2056/3946, 52.1%) perceived more difficulty in obtaining hand sanitizer compared to those in Taiwan (2708/63,634, 4.3%) and Hong Kong (104/1201, 8.7%). Only a small proportion of individuals in all countries (<5% each) reported having difficulty obtaining hand soap.

For cues to action, respondents selected up to three sources of information for COVID-19. News (eg, TV news, newspaper, and radio) and social media were the most frequently reported sources of information in every country, with Hong Kong reporting the highest percentage (news: 1002/1201, 83.4%; social media: 846/1201, 70.4%) and Mexico the lowest (news: 1965/3946, 49.8%; social media: 1337/3946, 33.9%). More respondents selected federal, or central, government rather than regional government as a top information source in Taiwan (14,730/63,634, 23.1% vs 1387/63,634, 2.2%) and Mexico (1155/3946, 29.3% vs 567/3946, 14.4%), while more respondents selected regional rather than federal government in the United States (1129/3070, 36.8% vs 312/3070, 10.2%). Both choices were comparably low in Hong Kong (<5% each).

Table 2. Weighted responses to health belief model (HBM) constructs by country.

HBM construct, survey question, and responses	Value (N=71,851), n (%)[a]				*P* value[b]
	United States (n=3070)	Mexico (n=3946)	Hong Kong (n=1201)	Taiwan (n=63,634)	
Perceived susceptibility to infection: What do you think is your risk of getting infected with COVID-19?					<.001
Not likely/slightly likely	1385 (45.1)	1253 (31.8)	635 (53.3)	44,345 (69.9)	—[c]
Moderately likely	1024 (33.4)	1523 (38.6)	331 (27.7)	10,038 (15.8)	—
Likely/very likely	659 (21.5)	1165 (29.6)	226 (19.0)	9042 (14.3)	—
Perceived severity of COVID-19: How afraid are you of the COVID-19 pandemic?					<.001
Not afraid/slightly afraid	876 (28.6)	1154 (29.3)	289 (24.2)	28,082 (44.2)	—
Moderately afraid	1022 (33.3)	1135 (28.8)	374 (31.3)	15,216 (24.0)	—
Afraid/very afraid	1166 (38.1)	1651 (41.9)	532 (44.5)	20,172 (31.8)	—
Perceived benefits of measures: How do you feel about the government measures for COVID-19 in your area?					<.001
Appropriate/essential	2062 (68.4)	1343 (34.7)	288 (24.5)	53,573 (85.5)	—
Unnecessarily restrictive/moderately restrictive	297 (9.9)	272 (7.0)	139 (11.9)	3227 (5.2)	—
Not enough	657 (21.8)	2256 (58.3)	745 (63.6)	5825 (9.3)	—
Self-efficacy in carrying out measures: How confident are you that you are able and willing to carry out these measures?					<.001
Not confident/slightly confident	138 (4.5)	93 (2.3)	52 (4.4)	3867 (6.1)	—
Moderately confident	216 (7.0)	248 (6.3)	195 (16.3)	11,470 (18.1)	—
Confident/very confident	2712 (88.5)	3599 (91.4)	946 (79.3)	48,144 (75.8)	—
Perceived barriers to carrying out measures: Have any barriers prevented you from adhering to these measures? Do you have any difficulty getting masks, hand soap, and hand sanitizer?					
Masks	1856 (60.5)	2547 (64.5)	633 (52.7)	7736 (12.2)	<.001
Essential job	381 (12.4)	630 (16.0)	333 (27.8)	16,141 (25.4)	<.001
Family obligations	201 (6.6)	636 (16.1)	92 (7.7)	5188 (8.2)	<.001
Transportation needs	49 (1.6)	193 (4.9)	595 (49.6)	15,158 (23.8)	<.001
Hand soap	142 (4.6)	167 (4.2)	13 (1.1)	384 (0.6)	<.001
Hand sanitizer	1536 (50.0)	2056 (52.1)	104 (8.7)	2708 (4.3)	<.001
Cues to action: What are your top three sources of information regarding COVID-19?[d]					
News source	2119 (69.0)	1965 (49.8)	1002 (83.4)	50,443 (79.3)	<.001
Social media	1234 (40.2)	1337 (33.9)	846 (70.4)	39,251 (61.7)	<.001
Central administration officials	312 (10.2)	1155 (29.3)	25 (2.1)	14,730 (23.1)	<.001
Regional administration officials	1129 (36.8)	567 (14.4)	60 (5.0)	1387 (2.2)	<.001

[a]Weighted values were calculated by dividing the actual proportion of the country's population by the proportion from the study's sample, then renormalized for each country to ensure weighted and unweighted sample sizes were equal. Due to rounding and missing data (<5% for each item), the sum of frequencies and percentages for the sample weighted columns may not equal the country's total sample size.

[b]*P* values were calculated using 2-sided chi-square tests.

[c]Not available.

[d]The top four media resources selected by respondents, when asked to pick their top three from the list, are shown.

Association of HBM Constructs With Handwashing and Social Distancing Behaviors

In multivariable analyses, individuals with higher self-efficacy were more likely to perform more handwashing practices compared to those with lower self-efficacy ($OR_{United\ States}$ 1.58, 95% CI 1.21-2.07; OR_{Mexico} 1.54, 95% CI 1.21-1.96; $OR_{Hong\ Kong}$ 2.48, 95% CI 1.80-3.44; OR_{Taiwan} 2.30, 95% CI 2.21-2.39) (see Table 3). The significance of other HBM constructs varied by country. Performing more handwashing practices was positively and significantly associated with perceived severity in the United States ($OR_{severity}$ 1.33, 95% CI 1.09-1.61), perceived susceptibility in Mexico ($OR_{susceptibility}$ 1.23, 95% CI 1.06-1.42) and Hong Kong ($OR_{susceptibility}$ 1.44, 95% CI 1.11-1.87), and perceived susceptibility and perceived severity in Taiwan ($OR_{susceptibility}$ 1.08, 95% CI 1.04-1.12; $OR_{severity}$ 1.24, 95% CI 1.20-1.29). In the United States, Mexico, and Taiwan, those who perceived government measures for COVID-19 as restrictive were significantly less likely to handwash compared to those who perceived measures as appropriate or essential ($OR_{United\ States}$ 0.41, 95% CI 0.30-0.55; OR_{Mexico} 0.65, 95% CI 0.50-0.86; OR_{Taiwan} 0.82, 95% CI 0.76-0.89). In the United States, non-White respondents were more likely to handwash compared to White or European respondents ($OR_{Hispanic/Latino\ or\ other}$ 3.22, 95% CI 2.69-3.86; OR_{Asian} 2.76, 95% CI 1.89-4.04). Similar patterns of association for covariates persisted even when we used handwashing time (>20 seconds), another important COVID-19 handwashing behavior, as the binary outcome in the sensitivity analysis (see Multimedia Appendix 5).

Similar to handwashing, individuals with higher self-efficacy were also more likely to practice social distancing compared to those with lower self-efficacy ($OR_{United\ States}$ 1.77, 95% CI 1.24-2.49; OR_{Mexico} 1.77, 95% CI 1.40-2.25; $OR_{Hong\ Kong}$ 3.25, 95% CI 2.32-4.62; OR_{Taiwan} 2.58, 95% CI 2.47-2.68) (see Table 4). Social distancing was also positively associated with perceived severity in the United States ($OR_{severity}$ 1.62, 95% CI 1.24-2.12), Mexico ($OR_{severity}$ 1.29, 95% CI 1.11-1.50), and Taiwan ($OR_{severity}$ 1.17, 95% CI 1.13-1.21). Similarly, in the United States, Mexico, and Taiwan, those who perceived government measures for COVID-19 as restrictive were significantly less likely to practice social distancing compared to those who perceived measures as appropriate or essential ($OR_{United\ States}$ 0.52, 95% CI 0.36-0.76; OR_{Mexico} 0.65, 95% CI 0.49-0.85; OR_{Taiwan} 0.82, 95% CI 0.76-0.88). In the United States, family obligations and transportation needs were associated with fewer social distancing practices (OR_{family} 0.25, 95% CI 0.17-0.36; $OR_{transportation}$ 0.25, 95% CI 0.11-0.57). In Hong Kong, those who had difficulty obtaining masks were more likely to socially distance (OR_{masks} 1.61, 95% CI 1.23-2.10), but those who had an essential job or transportation needs were less likely to socially distance ($OR_{essential\ job}$ 0.66, 95% CI 0.48-0.89; $OR_{transportation}$ 0.67, 95% CI 0.52-0.87). Similarly in Taiwan, those who had an essential job or transportation needs were less likely to socially distance ($OR_{essential\ job}$ 0.71, 95% CI 0.68-0.74; $OR_{transportation}$ 0.85, 95% CI 0.82-0.89).

Table 3. Multivariable model of health beliefs and handwashing practices by country.

Characteristic or construct and responses	United States, OR[a] (95% CI)	*P* value	Mexico, OR (95% CI)	*P* value	Hong Kong, OR (95% CI)	*P* value	Taiwan, OR (95% CI)	*P* value
Age group (years)								
18-24	Reference		Reference		Reference		Reference	
25-34	1.14 (0.65-2.03)	.70	1.15 (0.91-1.46)	.20	2.00 (1.14-3.55)	.02	1.44 (1.34-1.55)	<.001
35-44	1.24 (0.72-2.20)	.40	1.60 (1.25-2.04)	<.001	2.15 (1.24-3.80)	.007	2.24 (2.09-2.41)	<.001
45-59	1.91 (1.11-3.34)	.02	1.86 (1.47-2.35)	<.001	2.20 (1.29-3.82)	.004	2.66 (2.48-2.85)	<.001
60+	1.10 (0.63-1.95)	.70	1.76 (1.37-2.26)	<.001	1.16 (0.67-2.03)	.60	2.96 (2.76-3.18)	<.001
Gender								
Female	Reference		Reference		Reference		Reference	
Male	0.84 (0.71-1.00)	.05	1.00 (0.87-1.14)	.90	1.09 (0.85-1.40)	.50	0.69 (0.66-0.71)	<.001
Other[b]	1.52 (0.62-3.93)	.40	1.09 (0.56-2.11)	.80	1.73 (0.80-3.84)	.20	1.06 (0.96-1.17)	.02
Race or ethnicity								
White or European	Reference		N/A[c]		N/A		N/A	
Hispanic/Latino or other[d]	3.22 (2.69-3.86)	<.001	N/A	N/A	N/A	N/A	N/A	N/A
Asian	2.76 (1.89-4.04)	<.001	N/A	N/A	N/A	N/A	N/A	N/A
Education								
Below college	Reference		Reference		Reference		Reference	
College and above	0.99 (0.71-1.40)	.90	1.14 (0.95-1.36)	.20	1.13 (0.84-1.53)	.40	0.84 (0.80-0.88)	<.001
Perceived susceptibility of infection								
Not likely/slightly likely	Reference		Reference		Reference		Reference	
Moderately to very likely	1.12 (0.95-1.33)	.20	1.23 (1.06-1.42)	.006	1.44 (1.11-1.87)	.007	1.08 (1.04-1.12)	<.001
Perceived severity of COVID-19								
Not afraid/slightly afraid	Reference		Reference		Reference		Reference	
Moderately to very afraid	1.33 (1.09-1.61)	.005	1.06 (0.91-1.22)	.50	1.22 (0.90-1.65)	.20	1.24 (1.20-1.29)	<.001
Perceived benefits of handwashing measures								
Unnecessarily restrictive/moderately restrictive	0.41 (0.30-0.55)	<.001	0.65 (0.50-0.86)	.002	0.65 (0.41-1.02)	.06	0.82 (0.76-0.89)	<.001
Appropriate/essential	Reference		Reference		Reference		Reference	
Not enough	1.49 (1.23-1.82)	<.001	0.97 (0.84-1.12)	.70	0.77 (0.56-1.07)	.12	0.91 (0.86-0.96)	.001
Self-efficacy in carrying out handwashing measures								
Not confident/moderately confident	Reference		Reference		Reference		Reference	
Confident/very confident	1.58 (1.21-2.07)	<.001	1.54 (1.21-1.96)	<.001	2.48 (1.80-3.44)	<.001	2.30 (2.21-2.39)	<.001

Characteristic or construct and responses	United States, OR[a] (95% CI)	*P* value	Mexico, OR (95% CI)	*P* value	Hong Kong, OR (95% CI)	*P* value	Taiwan, OR (95% CI)	*P* value
Perceived barriers to following handwashing measures (reference is "no")								
Hand soap	0.73 (0.49-1.07)	.11	1.35 (0.98-1.87)	.07	7.59 (1.88-53.9)	.01	1.01 (0.81-1.27)	.90
Hand sanitizer	0.88 (0.74-1.03)	.11	1.01 (0.88-1.15)	.90	1.14 (0.74-1.77)	.50	0.86 (0.79-0.94)	<.001
Cues to action (reference is "no")[e]								
News source	0.77 (0.64-0.92)	.003	0.80 (0.70-0.92)	.001	0.97 (0.69-1.35)	.80	0.95 (0.91-0.99)	.02
Social media	0.60 (0.50-0.71)	<.001	0.77 (0.67-0.89)	<.001	0.79 (0.60-1.05)	.10	0.86 (0.83-0.89)	<.001
Central administration officials	1.32 (1.01-1.74)	.05	0.84 (0.73-0.98)	.02	1.27 (0.55-3.02)	.60	0.99 (0.95-1.03)	.70
Regional administration officials	0.63 (0.52-0.75)	<.001	0.83 (0.69-1.00)	.05	0.42 (0.22-0.77)	.007	1.08 (0.96-1.21)	.20

[a]OR: odds ratio; models were run using weighted data, which were calculated by dividing the actual proportion of the country's population by the proportion from the study's sample, then renormalized for each country to ensure weighted and unweighted sample sizes were equal.

[b]Responses of *other* gender include individuals who chose nonbinary/third gender, prefer not to say, or other (<3% of total responses).

[c]N/A: not applicable; race or ethnicity was not adjusted for these countries as the majority identified as the same race or ethnicity.

[d]Responses of *other* race or ethnicity include individuals who are Black or African American, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, or other. Categories were collapsed due to low numbers (<2% of total responses).

[e]The top four media resources selected by respondents, when asked to pick their top three from the list, are shown.

Table 4. Multivariable model of health beliefs and social distancing practices by country.

Characteristic or construct and responses	United States, OR (95% CI)[a]	*P* value	Mexico, OR (95% CI)	*P* value	Hong Kong, OR (95% CI)	*P* value	Taiwan, OR (95% CI)	*P* value
Age group (years)								
18-24	Reference		Reference		Reference		Reference	
25-34	1.21 (0.52-2.66)	.60	0.70 (0.54-0.90)	.006	1.19 (0.68-2.08)	.50	1.39 (1.29-1.50)	<.001
35-44	1.61 (0.71-3.53)	.20	0.69 (0.53-0.89)	.005	1.09 (0.63-1.89)	.80	1.82 (1.69-1.96)	<.001
45-59	1.67 (0.72-3.65)	.20	0.67 (0.52-0.86)	.002	1.17 (0.69-1.99)	.60	1.93 (1.80-2.07)	<.001
60+	3.30 (1.42-7.31)	.004	0.52 (0.40-0.68)	<.001	0.56 (0.32-0.97)	.04	1.89 (1.76-2.03)	<.001
Gender								
Female	Reference		Reference		Reference		Reference	
Male	1.36 (1.05-1.75)	.02	1.14 (0.99-1.31)	.07	1.04 (0.81-1.34)	.80	1.03 (1.00-1.07)	.07
Other[b]	1.68 (0.47-8.23)	.50	1.02 (0.52-2.06)	.90	0.38 (0.15-0.87)	.03	0.92 (0.83-1.01)	.10
Race or ethnicity								
White or European	Reference		N/A[c]		N/A		N/A	
Hispanic/Latino or other[d]	0.46 (0.35-0.61)	<.001	N/A	N/A	N/A	N/A	N/A	N/A
Asian	0.78 (0.44-1.42)	.40	N/A	N/A	N/A	N/A	N/A	N/A
Education								
Below college	Reference		Reference		Reference		Reference	
College and above	0.60 (0.36-0.96)	.04	1.32 (1.10-1.60)	.003	1.13 (0.83-1.54)	.40	1.17 (1.12-1.24)	<.001
Reduced income								
No	Reference		Reference		Reference		Reference	
Yes	0.60 (0.46-0.78)	<.001	0.95 (0.83-1.10)	.50	0.98 (0.74-1.30)	.90	1.07 (1.03-1.12)	.002
Perceived susceptibility of infection								
Not likely/slightly likely	Reference		Reference		Reference		Reference	
Moderately to very likely	1.11 (0.86-1.44)	.40	1.02 (0.88-1.19)	.80	0.76 (0.58-0.99)	.05	0.94 (0.90-0.97)	<.001
Perceived severity of COVID-19								
Not afraid/slightly afraid	Reference		Reference		Reference		Reference	
Moderately to very afraid	1.62 (1.24-2.12)	<.001	1.29 (1.11-1.50)	.001	1.34 (0.99-1.84)	.06	1.17 (1.13-1.21)	<.001
Perceived benefits of social distancing measures								
Unnecessarily restrictive/moderately restrictive	0.52 (0.36-0.76)	<.001	0.65 (0.49-0.85)	.002	1.24 (0.79-1.96)	.40	0.82 (0.76-0.88)	<.001
Appropriate/essential	Reference		Reference		Reference		Reference	
Not enough	1.72 (1.24-2.42)	.001	1.22 (1.05-1.41)	.01	1.04 (0.74-1.46)	.80	1.05 (0.99-1.11)	.13
Self-efficacy in carrying out social distancing measures								
Not confident/moderately confident	Reference		Reference		Reference		Reference	

Characteristic or construct and responses	United States, OR (95% CI)[a]	*P* value	Mexico, OR (95% CI)	*P* value	Hong Kong, OR (95% CI)	*P* value	Taiwan, OR (95% CI)	*P* value
Confident/very confident	1.77 (1.24-2.49)	.001	1.77 (1.40-2.25)	<.001	3.25 (2.32-4.62)	<.001	2.58 (2.47-2.68)	<.001
Perceived barriers to following social distancing measures (reference is "no")								
Masks	0.95 (0.73-1.23)	.70	1.11 (0.96-1.28)	.20	1.61 (1.23-2.10)	<.001	0.92 (0.88-0.97)	.002
Essential job	0.86 (0.61-1.23)	.40	0.85 (0.71-1.03)	.09	0.66 (0.48-0.89)	.007	0.71 (0.68-0.74)	<.001
Family obligations	0.25 (0.17-0.36)	<.001	0.84 (0.70-1.01)	.06	0.80 (0.50-1.29)	.40	0.97 (0.92-1.04)	.40
Transportation needs	0.25 (0.11-0.57)	<.001	0.78 (0.57-1.06)	.11	0.67 (0.52-0.87)	.002	0.85 (0.82-0.89)	<.001
Cues to action (reference is "no")[e]								
News source	1.58 (1.22-2.04)	<.001	0.96 (0.84-1.10)	.60	0.94 (0.66-1.33)	.70	0.93 (0.89-0.97)	<.001
Social media	0.53 (0.41-0.68)	<.001	0.92 (0.80-1.07)	.30	0.85 (0.64-1.13)	.30	0.91 (0.88-0.95)	<.001
Central administration officials	1.15 (0.78-1.75)	.50	1.04 (0.89-1.22)	.60	1.33 (0.56-3.13)	.50	1.15 (1.11-1.20)	<.001
Regional administration officials	0.70 (0.53-0.92)	.01	1.20 (0.98-1.46)	.08	1.59 (0.90-2.83)	.11	1.07 (0.96-1.20)	.20

[a]OR: odds ratio; models were run using weighted data, which were calculated by dividing the actual proportion of the country's population by the proportion from the study's sample, then renormalized for each country to ensure weighted and unweighted sample sizes were equal.

[b]Responses of *other* gender include individuals who chose nonbinary/third gender, prefer not to say, or other (<3% of total responses).

[c]N/A: not applicable; race or ethnicity was not adjusted for these countries as the majority identified as the same race or ethnicity.

[d]Responses of *other* race or ethnicity include individuals who are Black or African American, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, or other. Categories were collapsed due to low numbers (<2% of total responses).

[e]The top four media resources selected by respondents, when asked to pick their top three from the list, are shown.

Discussion

Principal Findings

In this international study to examine COVID-19-related health behaviors using the HBM, we showed that respondents from the United States practiced the most social distancing, while those from Taiwan practiced the most handwashing. Despite these differences in health behaviors, self-efficacy was a significant predictor in all four countries. Our findings may be explained by the strictness and timing of government policies, the number of confirmed infection cases in each country, individual beliefs, and cultural context.

In the context of government interventions, Taiwan's early border control, case identification, isolation of suspected cases, and resource allocation led to recommendations for social distancing, though not strictly enforced [27]. Similarly, Hong Kong's early identification and strict quarantine of suspected cases resulted in regulations that prohibited large public gatherings but otherwise maintained regular activities [28,29]. On the other hand, 43 US states issued lockdown orders between mid-March and early April 2020, each lasting until the end of April at the minimum [30]. Mexico issued similar orders on March 26, 2020,with the strictest measures lasting until the end of May 2020 [31]. These varying degrees of strictness and timing of government interventions among the four countries may have attributed to increased social distancing in the United States and Mexico compared to Hong Kong and Taiwan.

Furthermore, at the start of our study period on April 11, 2020, the World Health Organization Situation Report recorded 1.6 million confirmed COVID-19 cases, with more than 99,000 deaths in over 200 countries and territories [32]. This included 461,275 confirmed cases in the United States, 3441 in Mexico, 1001 in Hong Kong, and 382 in Taiwan [26,32,33]. By the end of our study period on May 1, 2020, the numbers of cases and deaths had doubled worldwide [32]. The United States and Mexico saw a 124% and 417% increase in the number of confirmed COVID-19 cases, respectively, while Taiwan and Hong Kong reported only a 6.8% and 3.9% increase, respectively [26,32,33]. The rapid increase in confirmed cases in the United States and Mexico compared to Hong Kong and Taiwan likely also played a role in understanding handwashing and social distancing behaviors.

Among the HBM constructs, our study found that self-efficacy was the strongest positive predictor for both handwashing and social distancing in all countries. These findings were largely consistent with previous studies that examined preventative

behaviors for cancers using the HBM [34,35]. Although a review of HBM studies suggested that the construct of perceived barriers was the best individual predictor across different types of studies and behaviors, self-efficacy can be seen as an important factor in overcoming the barriers to taking actions [17]. Perceived barriers were not significantly associated with hand hygiene, possibly because difficulties accessing water, soap, and hand sanitizer were less common among our survey respondents. For instance, Mexico is an upper-middle-income country and in some regions and communities these items might not be readily available; however, the population we reached through social media may be comparable to the populations we assessed from the other high-income countries [36]. For social distancing behaviors, having transportation needs was consistently associated with practicing less social distancing. This is especially relevant in Hong Kong and Taiwan, where public transportation is heavily utilized with their population densities of 6690 and 652 persons per square kilometer, respectively [37,38]. In their most densely populated districts, these numbers are even higher at 57,250 and 27,418 persons per square kilometer, respectively [37,39]. It is also important to note that we treated each perceived barrier as a unique covariate in the model to assess the most relevant barriers to social distancing, which may differ from other studies. Moreover, given that our respondents were mostly well-educated social media users, the true proportion of individuals with perceived barriers in our study was likely underestimated.

Previous studies have also suggested perceived susceptibility to be a good predictor for preventative behaviors [17]. In our study, perceived susceptibility was, overall, a significant positive predictor for practicing more handwashing. Perceived severity was also a strong predictor for both handwashing and social distancing in the United States and Taiwan, which may have been influenced by the worldwide news coverage and the strictness of government interventions. The associations found between perceived benefits and health behaviors may be tied to the timing of policies in each country and overall trust in the government. Cues to action, measured as types of media publicity, were not significant for predicting behavioral change in our model.

Modifiable factors that influence individual beliefs, such as culture and prior knowledge, are important to consider. In Hong Kong and Taiwan, wide adoption of preventative behaviors after the 2003 SARS outbreak may have better prepared residents for COVID-19, which may explain their greater sense of self-efficacy in handwashing and social distancing compared to other countries. Many residents were already taking regular individual actions, practicing good hand hygiene for infection control or wearing masks to counter air pollution when the pandemic hit. In fact, the study team received several emails from respondents in Taiwan and Hong Kong, noting that they had practiced handwashing prior to the pandemic because they were taught to do so as children. For this study, we were unable to statistically account for social factors and prior knowledge in our analyses, but future studies should consider including them into models to assess the influence of social and cultural factors on preventative health behaviors.

Using Social Media for Recruitment During COVID-19

Our study may also provide insight into the effect of using social media recruitment strategies to reach a large audience. Given the rapidly evolving information, beliefs, and policies surrounding COVID-19, internet sampling allowed us to (1) capture real-time data simultaneously in different countries in a short time span, (2) reach a large number of participants in lockdown, and (3) overcome financial limitations [40]. The combination of boosting and sharing of social media posts allowed us to effectively target specific populations and locations while also reaching a larger audience, as was evident by the number of respondents from Taiwan. Our findings expand on recent COVID-19 studies from the United States and Taiwan that used similar methods to assess other COVID-19 attitudes, behaviors, and knowledge among different populations [41-43].

Limitations

There are limitations to this study. Firstly, we used convenience sampling to recruit participants, which could have introduced potential sample selection bias. For example, we found an underrepresentation of populations with lower educational levels. This may have resulted in an overestimation of adherence rates and underestimation of perceived barriers. However, in multivariate analysis, education was not statistically associated with handwashing or social distancing practices. To best address the imbalances in our sample, we conducted poststratification weighting by age and gender, as well as race or ethnicity for the United States, to improve the generalizability of our results, although we understand that this does not make up for all of the differences [44]. Secondly, we had a disproportionately larger sample size in Taiwan relative to other countries [45]. However, since our main multivariable analyses were country specific, this would not likely affect the estimates found in other countries. Finally, there are weaknesses within the HBM itself. The HBM does not account for a person's nonhealth-related beliefs or determinants that dictate a person's acceptance of a health behavior. Health behaviors can also be learned through modeling as explained by other behavior change theories, such as the social cognitive theory (SCT); for instance, residents in Taiwan and Hong Kong might regularly wear masks and practice hand hygiene from observing those around them [46]. Self-efficacy, the strongest predictor in our study, is also known to play a large role in health behavior in the context of the SCT. However, we did not use SCT in our study because the theory's heavy emphasis on the process of learning disregards an individual's perception about COVID-19 as well as their motivations behind handwashing and social distancing behaviors. The collection of data on health beliefs, which the HBM encompasses, is important for the planning of interventions that can then be targeted to each country's specific needs.

Conclusions

Overall, our findings revealed that certain health belief constructs were independently associated with social distancing and handwashing behaviors. In the context of controlling the continued spread of COVID-19, self-efficacy is a significant predictor that can be easily targeted and modified by public health officials and educators. Policies and communications

that address relevant health beliefs can facilitate adoption of necessary actions for preventing COVID-19.

Acknowledgments

We would like to thank Beth Duff-Brown for disseminating the survey via social media and Judy Sun for helping with the Mandarin translation of study-related materials.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey items.
[DOCX File , 21 KB - jmir_v23i2e23720_app1.docx]

Multimedia Appendix 2
Comparison of unweighted and weighted sample characteristics in the United States and Mexico relative to country population estimates.
[DOCX File , 22 KB - jmir_v23i2e23720_app2.docx]

Multimedia Appendix 3
Comparison of unweighted and weighted sample characteristics in Hong Kong and Taiwan relative to country population estimates.
[DOCX File , 22 KB - jmir_v23i2e23720_app3.docx]

Multimedia Appendix 4
Multivariable models assessing handwashing and social distancing practices by country.
[DOCX File , 16 KB - jmir_v23i2e23720_app4.docx]

Multimedia Appendix 5
Multivariable models of health beliefs and handwashing time (>20 seconds) by country.
[DOCX File , 19 KB - jmir_v23i2e23720_app5.docx]

References

1. COVID-19 Strategy Update. Geneva, Switzerland: World Health Organization; 2020 Apr 14. URL: https://www.who.int/docs/default-source/coronaviruse/covid-strategy-update-14april2020.pdf [accessed 2020-06-14]
2. COVID-19 impact on education. UNESCO. 2020. URL: https://en.unesco.org/covid19/educationresponse [accessed 2020-06-14]
3. Rashid H, Ridda I, King C, Begun M, Tekin H, Wood JG, et al. Evidence compendium and advice on social distancing and other related measures for response to an influenza pandemic. Paediatr Respir Rev 2015 Mar;16(2):119-126. [doi: 10.1016/j.prrv.2014.01.003] [Medline: 24630149]
4. Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature 2020 Jun 08;584(7820):262-267. [doi: 10.1038/s41586-020-2404-8] [Medline: 32512578]
5. Blendon RJ, Koonin LM, Benson JM, Cetron MS, Pollard WE, Mitchell EW, et al. Public response to community mitigation measures for pandemic influenza. Emerg Infect Dis 2008 May;14(5):778-786 [FREE Full text] [doi: 10.3201/eid1405.071437] [Medline: 18439361]
6. Jarvis CI, Van Zandvoort K, Gimma A, Prem K, CMMID COVID-19 Working Group, Klepac P, et al. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. BMC Med 2020 May 07;18(1):124 [FREE Full text] [doi: 10.1186/s12916-020-01597-8] [Medline: 32375776]
7. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: A quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]
8. Briscese G, Lacetera N, Macis M, Tonin M. Expectations, Reference Points, and Compliance With COVID-19 Social Distancing Measures. Cambridge, MA: National Bureau of Economic Research; 2020 Mar. URL: https://www.nber.org/system/files/working_papers/w26916/w26916.pdf [accessed 2021-02-21]
9. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]

10. Digital 2020: The United States of America. DataReportal. 2020 Feb 11. URL: https://datareportal.com/reports/digital-2020-united-states-of-america [accessed 2020-08-21]
11. Digital 2020: Mexico. DataReportal. 2020 Feb 11. URL: https://datareportal.com/reports/digital-2020-mexico [accessed 2020-08-21]
12. Digital 2020: Hong Kong. DataReportal. 2020 Feb 13. URL: https://datareportal.com/reports/digital-2020-hong-kong [accessed 2020-08-21]
13. Digital 2020: Taiwan. DataReportal. 2020 Feb 18. URL: https://datareportal.com/reports/digital-2020-taiwan [accessed 2020-08-21]
14. Leiner DJ. Our research's breadth lives on convenience samples: A case study of the online respondent pool "SoSci Panel". Stud Commun Media 2016 Apr;5(4):367-396 [FREE Full text] [doi: 10.5771/2192-4007-2016-4-367]
15. Kennedy C, Hartig H. Response rates in telephone surveys have resumed their decline. Pew Research Center. Washington, DC: Pew Research Center; 2019 Feb 27. URL: https://www.pewresearch.org/fact-tank/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/ [accessed 2020-06-14]
16. Lavrakas PJ. Surveys by telephone. In: Donsbach W, Traugott MW, editors. The SAGE Handbook of Public Opinion Research. London, UK: SAGE Publications Ltd; 2008:249-261.
17. Champion VL, Skinner CS. The health belief model. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. San Francisco, CA: Jossey-Bass; 2008:45-65.
18. ACS demographic and housing estimates. United States Census Bureau. 2018. URL: https://data.census.gov/cedsci/table?q=race&hidePreview=false&tid=ACSDP1Y2018.DP05&vintage=2018 [accessed 2020-06-24]
19. International Data Base (IDB). United States Census Bureau. URL: https://www.census.gov/data-tools/demo/idb/#/country?YR_ANIM=2020&FIPS_SINGLE= MX&dashPages=BYAGE&COUNTRY_YEAR=2020&ageGroup=5Y [accessed 2021-02-22]
20. Hong Kong Census and Statistics Department. Main tables. 2016 Population By-census. 2016. URL: https://www.bycensus2016.gov.hk/en/bc-mt.html [accessed 2020-06-24]
21. 2010 population and housing census. The Directorate General of Budget, Accounting and Statistics, Executive Yuan, Republic of China (Taiwan). 2010. URL: https://census.dgbas.gov.tw/PHC2010/english/rehome.htm [accessed 2020-06-24]
22. Royal KD. Survey research methods: A guide for creating post-stratification weights to correct for sample bias. Educ Health Prof 2019;2(1):48-50. [doi: 10.4103/ehp.ehp_8_19]
23. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html [accessed 2020-06-24]
24. Quédate en casa. Gobierno de México. 2020. URL: https://coronavirus.gob.mx/quedate-en-casa/ [accessed 2020-06-24]
25. Health advice. The Government of the Hong Kong Special Administrative Region. 2020. URL: https://www.coronavirus.gov.hk/eng/health-advice.html [accessed 2020-06-24]
26. Taiwan Centers for Disease Control. 2020. URL: https://www.cdc.gov.tw/En [accessed 2020-06-14]
27. Wang CJ, Ng CY, Brook RH. Response to COVID-19 in Taiwan: Big data analytics, new technology, and proactive testing. JAMA 2020 Apr 14;323(14):1341-1342. [doi: 10.1001/jama.2020.3151] [Medline: 32125371]
28. Cheng VCC, Wong S, Chen JHK, Yip CCY, Chuang VWM, Tsang OTY, et al. Escalating infection control response to the rapidly evolving epidemiology of the coronavirus disease 2019 (COVID-19) due to SARS-CoV-2 in Hong Kong. Infect Control Hosp Epidemiol 2020 May;41(5):493-498 [FREE Full text] [doi: 10.1017/ice.2020.58] [Medline: 32131908]
29. New requirements to reduce gatherings. The Government of the Hong Kong Special Administrative Region. 2020. URL: https://www.coronavirus.gov.hk/eng/reduce-gatherings.html [accessed 2020-06-14]
30. COVID-19 policy tracker. MultiState. 2020. URL: https://www.multistate.us/pages/covid-19-policy-tracker [accessed 2020-06-14]
31. COVID-19 information for US citizens in Mexico. US Embassy & Consulates in Mexico. 2020. URL: https://mx.usembassy.gov/u-s-citizen-services/covid-19-information/ [accessed 2020-06-14]
32. Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update. World Health Organization. 2020. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports [accessed 2020-06-14]
33. Latest situation of cases of COVID-19 (as of 14 June 2020). Centre for Health Protection, The Government of the Hong Kong Special Administrative Region. 2020 Jun. URL: https://www.chp.gov.hk/files/pdf/local_situation_covid19_en.pdf [accessed 2020-06-14]
34. Dewi T, Massar K, Ruiter R, Leonardi T. Determinants of breast self-examination practice among women in Surabaya, Indonesia: An application of the health belief model. BMC Public Health 2019 Nov 27;19(1):1581 [FREE Full text] [doi: 10.1186/s12889-019-7951-2] [Medline: 31775697]
35. Kasparian NA, Bränström R, Chang Y, Affleck P, Aspinwall LG, Tibben A, Melanoma Genetics Consortium (GenoMEL). Skin examination behavior: The role of melanoma history, skin type, psychosocial factors, and region of residence in determining clinical and self-conducted skin examination. Arch Dermatol 2012 Oct;148(10):1142-1151 [FREE Full text] [doi: 10.1001/archdermatol.2012.1817] [Medline: 22801744]
36. World Bank country and lending groups. The World Bank. 2020. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [accessed 2020-06-30]

37. Hong Kong: The Facts - Population. The Hong Kong Special Administrative Region Government. 2015 Apr. URL: https://www.gov.hk/en/about/abouthk/factsheets/docs/population.pdf [accessed 2020-06-21]
38. Latest indicators. National Statistics, Republic of China (Taiwan). 2020. URL: https://eng.stat.gov.tw/point.asp?index=9 [accessed 2020-06-21]
39. Demographic overview. Taipei City Government. 2017. URL: https://english.gov.taipei/cp.aspx?n=C619997124A6D293 [accessed 2020-06-21]
40. Ali SH, Foreman J, Capasso A, Jones AM, Tozan Y, DiClemente RJ. Social media as a recruitment platform for a nationwide online survey of COVID-19 knowledge, beliefs, and practices in the United States: Methodology and feasibility analysis. BMC Med Res Methodol 2020 May 13;20(1):116 [FREE Full text] [doi: 10.1186/s12874-020-01011-0] [Medline: 32404050]
41. Oosterhoff B, Palmer CA. Attitudes and psychological factors associated with news monitoring, social distancing, disinfecting, and hoarding behaviors among US adolescents during the coronavirus disease 2019 pandemic. JAMA Pediatr 2020 Jun 29;174(12):1184-1190. [doi: 10.1001/jamapediatrics.2020.1876] [Medline: 32597925]
42. Clements JM. Knowledge and behaviors toward COVID-19 among US residents during the early days of the pandemic: Cross-sectional online questionnaire. JMIR Public Health Surveill 2020 May 08;6(2):e19161 [FREE Full text] [doi: 10.2196/19161] [Medline: 32369759]
43. Wang P, Lu W, Ko N, Chen Y, Li D, Chang Y, et al. COVID-19-related information sources and the relationship with confidence in people coping with COVID-19: Facebook survey study in Taiwan. J Med Internet Res 2020 Jun 05;22(6):e20021 [FREE Full text] [doi: 10.2196/20021] [Medline: 32490839]
44. Franco A, Malhotra N, Simonovits G, Zigerell LJ. Developing standards for post-hoc weighting in population-based survey experiments. J Exp Pol Sci 2017 Oct 12;4(2):161-172. [doi: 10.1017/xps.2017.2]
45. Faber J, Fonseca LM. How sample size influences research outcomes. Dental Press J Orthod 2014;19(4):27-29 [FREE Full text] [doi: 10.1590/2176-9451.19.4.027-029.ebo] [Medline: 25279518]
46. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [doi: 10.1037/0003-066x.44.9.1175] [Medline: 2782727]

Abbreviations

HBM: health belief model
OR: odds ratio
SCT: social cognitive theory

Edited by G Fagherazzi; submitted 21.08.20; peer-reviewed by CF Yen, D Goyal; comments to author 09.12.20; revised version received 16.12.20; accepted 01.02.21; published 26.02.21.

Please cite as:
Hsing JC, Ma J, Barrero-Castillero A, Jani SG, Pulendran UP, Lin BJ, Thomas-Uribe M, Wang CJ
Influence of Health Beliefs on Adherence to COVID-19 Preventative Practices: International, Social Media–Based Survey Study
J Med Internet Res 2021;23(2):e23720
URL: https://www.jmir.org/2021/2/e23720
doi:10.2196/23720
PMID:33571103

Original Paper

Novel Analgesic Index for Postoperative Pain Assessment Based on a Photoplethysmographic Spectrogram and Convolutional Neural Network: Observational Study

Byung-Moon Choi[1*], MD, PhD; Ji Yeon Yim[2*], BSc; Hangsik Shin[2*], PhD; Gyujeong Noh[1,3], MD, PhD

[1]Department of Anaesthesiology and Pain Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[2]Department of Biomedical Engineering, Chonnam National University, Yeosu, Republic of Korea
[3]Department of Clinical Pharmacology and Therapeutics, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Hangsik Shin, PhD
Department of Biomedical Engineering
Chonnam National University
50 Daehak-ro
Yeosu, 59626
Republic of Korea
Phone: 82 1068998679
Email: hangsik.shin@jnu.ac.kr

Abstract

Background: Although commercially available analgesic indices based on biosignal processing have been used to quantify nociception during general anesthesia, their performance is low in conscious patients. Therefore, there is a need to develop a new analgesic index with improved performance to quantify postoperative pain in conscious patients.

Objective: This study aimed to develop a new analgesic index using photoplethysmogram (PPG) spectrograms and a convolutional neural network (CNN) to objectively assess pain in conscious patients.

Methods: PPGs were obtained from a group of surgical patients for 6 minutes both in the absence (preoperatively) and in the presence (postoperatively) of pain. Then, the PPG data of the latter 5 minutes were used for analysis. Based on the PPGs and a CNN, we developed a spectrogram–CNN index for pain assessment. The area under the curve (AUC) of the receiver-operating characteristic curve was measured to evaluate the performance of the 2 indices.

Results: PPGs from 100 patients were used to develop the spectrogram–CNN index. When there was pain, the mean (95% CI) spectrogram–CNN index value increased significantly—baseline: 28.5 (24.2-30.7) versus recovery area: 65.7 (60.5-68.3); $P<.01$. The AUC and balanced accuracy were 0.76 and 71.4%, respectively. The spectrogram–CNN index cutoff value for detecting pain was 48, with a sensitivity of 68.3% and specificity of 73.8%.

Conclusions: Although there were limitations to the study design, we confirmed that the spectrogram–CNN index can efficiently detect postoperative pain in conscious patients. Further studies are required to assess the spectrogram–CNN index's feasibility and prevent overfitting to various populations, including patients under general anesthesia.

Trial Registration: Clinical Research Information Service KCT0002080; https://cris.nih.go.kr/cris/search/search_result_st01.jsp?seq=6638

(J Med Internet Res 2021;23(2):e23920) doi:10.2196/23920

KEYWORDS

analgesic index; machine learning; pain assessment; photoplethysmogram; postoperative pain; spectrogram

Introduction

Efficient management of postoperative pain affecting the prognosis of patients is becoming increasingly important [1]. To properly administer analgesics, it is necessary to first objectively assess the patient's degree of pain. In conscious patients, pain can be assessed by asking the patient directly, but unconscious patients or those with difficulty communicating

require an appropriate index to quantify their pain. However, current commercial analgesic indices were developed for the purpose of evaluating nociception in patients under general anesthesia [2,3]; therefore, there is no standard for the quantification of postoperative pain in conscious patients [4]. Thus, developing a new pain index to quantify pain in patients who cannot directly communicate their level of pain may also help in the clinical setting; it will also reduce the need to ask questions each time the patient is conscious when pain must be evaluated frequently.

A photoplethysmogram (PPG) is a biosignal that can be obtained continuously and noninvasively using a pulse oximeter. Because a PPG conveys much information about a patient's condition, many attempts have been made to quantify pain by analyzing PPG signals [3,5-7]. The surgical pleth index (SPI; GE Healthcare), developed for quantifying nociception during general anesthesia, only considers the amplitude and heartbeat interval of a PPG [3]. In addition to these 2 parameters, other pain-related features are present in PPG signals [6,7]. Therefore, the application of a new analytical method has the potential to improve the performance of analgesic indices.

Deep learning architectures, such as a convolution neural network (CNN), can be a good solution to elucidate the hidden features in a PPG because they can identify optimal abstracted features that are beyond human comprehension without any manual procedure [8]. Furthermore, in determining the presence of pain, machine learning has a strong advantage owing to its nonlinear characteristics compared with the SPI, which assesses pain based on simple linear regression [8,9], potentially making it possible to effectively predict nonlinear deviations among individuals or situations [10]. Therefore, a combination of the extended features of PPG and machine learning–based scoring is expected to overcome the limitations of existing pain assessment techniques. However, because a PPG is a 1D signal, whereas CNNs have the advantage of multidimensional data analysis, a dimensional extension of a PPG without loss of time–frequency characteristics is required to apply it optimally in a CNN. A spectrogram, which is a 2D image including the intact time–frequency characteristics of a PPG, can be a good method for applying a CNN to PPGs.

This study aimed to develop a new analgesic index using PPG spectrograms and a CNN to objectively assess pain in conscious patients. In addition, the performance of our newly developed index was compared with that of the SPI.

Methods

Patient Population

The study protocol was approved by the Institutional Review Board of Asan Medical Centre (approval number: 2016-0477) and registered on an international clinical trials registry platform (registration number KCT0002080). Written informed consent was obtained from all patients. All procedures were conducted in accordance with relevant guidelines and regulations. In total, 120 patients (American Society of Anesthesiologists Physical Status 1, 2, or 3) between the ages of 20 and 80, who were scheduled to undergo elective surgery, were included in this observational study. Exclusion criteria were as follows: clinically significant impairment of the cardiovascular, hepatic, or renal function; history of cardiac arrhythmia; use of medication that might affect autonomic function; the presence of presurgical acute or chronic pain (Visual Analog Scale score [VAS] > 0, measured before surgery); clinically significant laboratory findings; and evidence of pregnancy.

Procedure and Data Acquisition

All patients fasted from midnight on the day of surgery without premedication. In the operating theater, patients were monitored for their heart activity using electrocardiography, end-tidal carbon dioxide partial pressure, and noninvasive blood pressure measurement. Neuromuscular transmission was monitored using an M-NMT module at the adductor pollicis muscle (CARESCAPE B850; GE Healthcare). A reusable SPI sensor was placed on the index finger of each patient (on the arm not used for blood pressure measurement). Patients were allowed to acclimatize for at least 5 minutes in the supine position in a quiet operating theater, after which baseline data (without pain) were collected for 6 minutes, of which the latter 5 minutes were used for analysis. General anesthesia was performed by administering propofol and remifentanil by a target effect-site concentration–controlled infusion using the Schnider and Minto models [11,12]. Target effect-site concentrations (*Ces*) of propofol were titrated to maintain the bispectral index (Covidien) at less than 60 during the induction and maintenance of anesthesia. The target *Ces* of remifentanil were adjusted to maintain stable hemodynamics (ie, systolic blood pressure >80 mmHg and heart rate over 45 beats/min). All patients received a bolus dose of oxycodone (0.1 mg/kg) 30 minutes prior to the end of surgery.

Intravenous patient-controlled analgesia with oxycodone began after the administration of the bolus dose of oxycodone. Neuromuscular blockade was reversed with neostigmine and glycopyrrolate at the end of surgery. Tracheal extubation was performed when the train-of-four ratio was greater than 0.9 and bispectral index value was greater than 80. Patients were then transported to the postanesthesia care unit (PACU). When the patients arrived in the PACU, their state of consciousness was assessed with a modified Aldrete score [13]. Electrocardiogram, pulse oximetry, and noninvasive blood pressure were also monitored. Additional PPG and SPI data were obtained for the initial 5 minutes in the PACU. After obtaining the data, patients were assessed for pain using a VAS (0=no pain; 100=the most severe pain). Oxycodone was administered according to postoperative pain intensity. The PPG and SPI values were measured using an S/5 Anesthesia Monitor (Datex-Ohmeda, Inc.) and recorded on a laptop for offline analysis. The PPG data were sampled at 300 Hz, and SPI data were recorded every 10 seconds.

Pain Assessment Model

A spectrogram–CNN model was developed and validated through fivefold cross-validation. The developed model outputs the spectrogram–CNN index as a pain score using a PPG spectrogram as input and CNN as a pain scorer. During model development, patients and test sets were separated to prevent intrasubject interference to avoid data overlaps between the

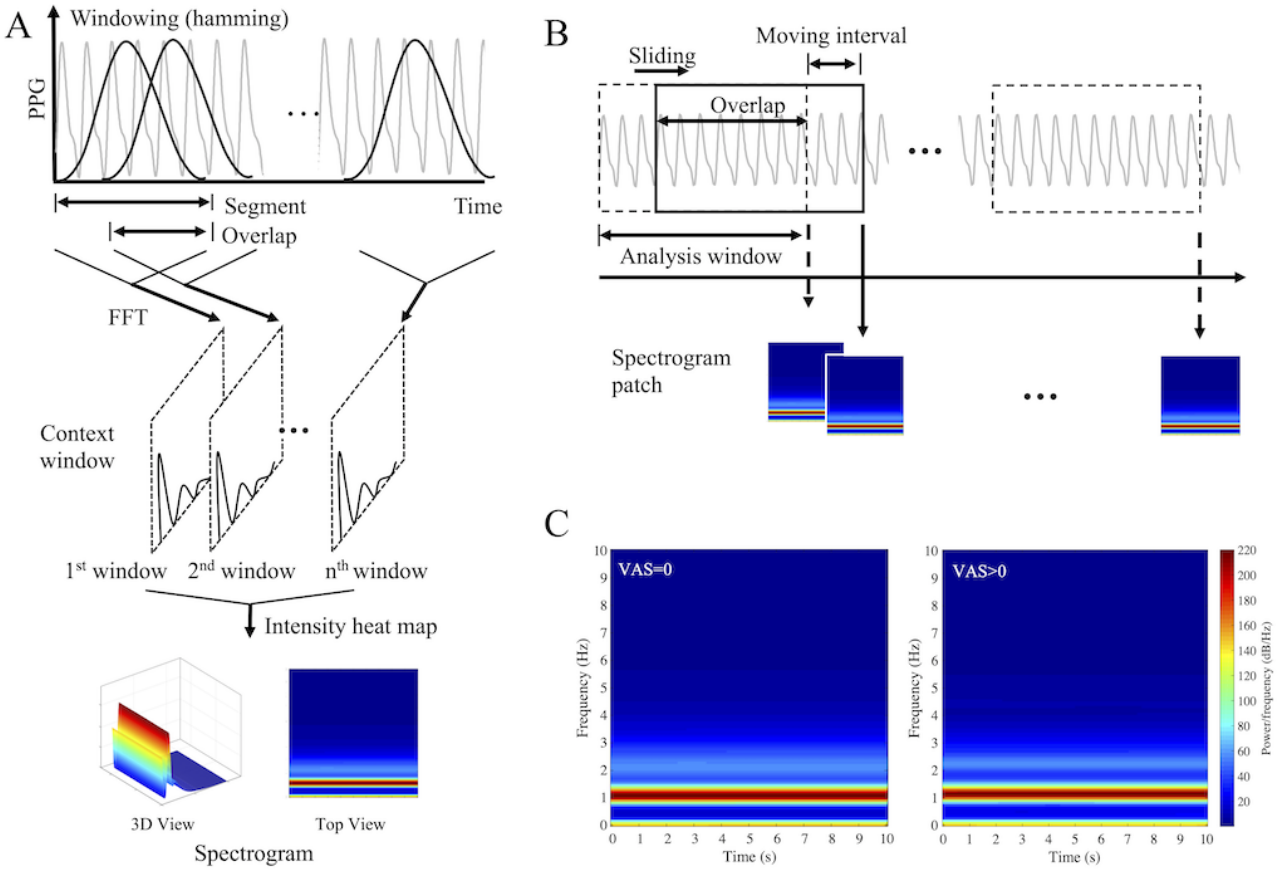
development and test sets of each fold. For model training, 90% of the development set was used as a training set and 10% as a validation set. Finally, for each fold, 20, 73, and 7 patients' data were used as a test set, training set, and validation set, respectively. Pain and nonpain data labels were created based on the VAS, where VAS>0 was defined as pain and VAS=0 was defined as nonpain with labels "1" and "0," respectively. Detailed descriptions of the spectrograms and CNN used in this study are given below.

Spectrogram

Spectrogram creation is a method used for time–frequency analysis of time series signals. A spectrogram reconstructs 2D images while maintaining the information contained in 1D time series data [14]. Spectrograms are useful for visually describing changes in the frequency characteristics of nonstationary signals, such as physiological signals over time [15,16]. Spectrograms can be generated by repeating short-time Fourier transforms that divide a longer time signal into shorter segments of equal length and then computing the Fourier transforms separately on each shorter segment. In this study, 2D spectrogram images generated from 1D PPGs were used as the pain classifier input to reflect the whole waveform and change of waveform, not the specific feature of the PPG waveform. Prior to spectrogram generation, all PPGs were filtered using both a finite impulse response bandpass filter with a 0.5-10-Hz passband and a 30-tap moving average filter. In addition, considering that PPG amplitude is an arbitrary unit, a normalization process was performed to reduce the intersubject and intermeasurement deviations [17]. In the normalization process, *z*-scores were obtained by subtracting the mean of the measured values from each measured value and dividing by the standard deviation. The spectrogram generation process is illustrated in Figure 1.

Spectrogram images were generated by short-time Fourier transforms of 10-second PPGs every 10 seconds without overlap (Figure 1A). At the time, to generate a single spectrogram image, each 10-second PPG was divided into 6.3-second segments with a 6.27-second overlap and transformed to the frequency domain using fast Fourier transform after windowing with a Hamming window (Figure 1B). The frequency range of the spectrogram image was set to 0-10 Hz, and the frequency resolution was set to 0.81 Hz to equalize the number of time frames and frequency bins, that is, to equalize the numbers of horizontal and vertical pixels in the spectrogram image, respectively. Finally, 30 spectrogram patches of size 124 × 124 were generated for each 5-minute PPG. Figure 1C shows the averaged spectrograms without pain (left) and with pain (right). All preprocessing and spectrogram patches were generated using MATLAB (version 2018a; The MathWorks, Inc.).

Figure 1. Process for generation of a single spectrogram (A), process of generating multiple spectrograms over time using a sliding window (B) and the average spectrogram during pain (left) and non-pain (right) conditions (C); FFT: fast Fourier transform; PPG: photoplethysmogram.

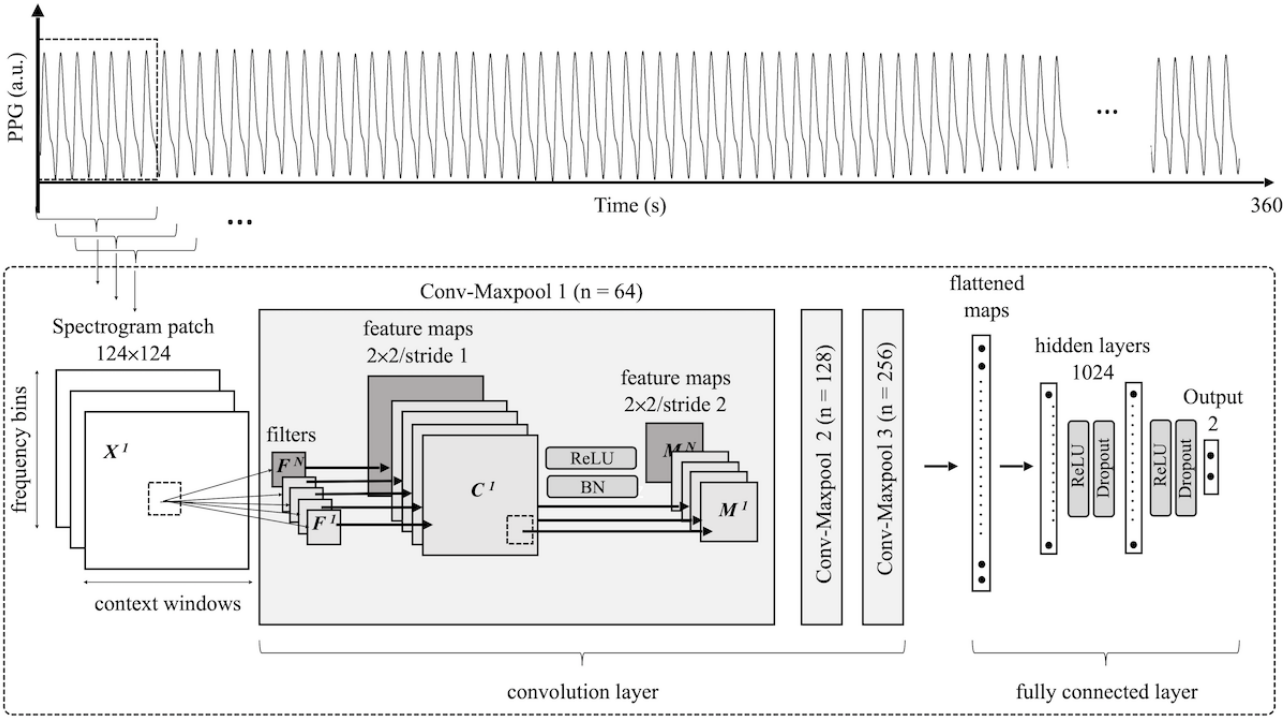


Convolutional Neural Network

CNNs are useful for image analysis because they have the advantage of maintaining the spatial information of 2D or higher inputs. CNNs enable data-driven learning, are highly representative, and effectively combine the spatial information of multidimensional inputs [18]; supervised CNNs extract information more effectively due to class-specific information

[19,20]. The CNN used in this study has a 2D PPG spectrogram input and binary-coded labels: “0” is a pain-free state and “1” is pain state. Figure 2 shows the structure of the CNN developed in this study. First, in the convolution-max pooling layer (Conv-Maxpool), 32, 64, and 128 filters are applied to the spectrogram input to perform a convolution process, and the spatial characteristics are simplified through max pooling. In Conv-Maxpool, the filter size of all convolutional layers is 2 × 2 and the stride is 1. The size of the feature map is reduced by setting the filter size of the max pooling layer to 2 × 2 and stride to 2, equal to the filter size. A batch normalization layer and rectified linear unit (ReLU) activation functions are applied to the Conv-Maxpool to increase learning speed and efficiency. Batch normalization is a structure that improves the speed and stability of neural networks by normalizing interlayer input data [21], and ReLU improves the expressive power of neural networks based on nonlinear features [22,23]. The Conv-Maxpool process is repeated 3 times, resulting in a feature map size of 15 × 15. The fully connected layer consists of 2 hidden layers and 1 output layer, and the ReLU activation function and dropout are also applied to the hidden layer to reduce overfitting [24]. The dropout rate was set to 0.5 in training, but there was no dropout during testing. Cross-entropy is employed as a cost function [25], and adaptive moment estimation (Adam) is used as an optimizer [26]. Finally, the result is output to 2 nodes, representing “pain free” and “pain,” and the values are probabilistically expressed using the SoftMax function [27]. Because the output of SoftMax gives the probability of the input data being a painful condition with a value between 0 and 1, it is converted into a pain index as the “likelihood of pain.” Consequently, the spectrogram–CNN index is calculated by multiplying the probability value output from the pain node by 100. The CNN model was implemented and trained using Python 3.7 (Python Software Foundation) and TensorFlow 2.0 in the Anaconda environment.

Figure 2. Architecture of the convolutional neural network proposed in this study. X: input, F: filter, C: convolution layer, M: max pooling, N: number of filters, BN: batch normalisation, ReLU: rectifier linear unit, Conv-Maxpool: convolution-max pooling.



Statistical Analyses

Receiver-operating characteristic (ROC) curves were computed to compare the sensitivity and specificity of the spectrogram–CNN index and SPI for detecting pain. Cutoff values used for calculation of sensitivity and specificity were computed as “best fit” (highest combined sensitivity and specificity) [28]. The differences between the spectrogram–CNN index and SPI in the ROC curves were calculated using the MedCalc Statistical Software (version 13.3.1; MedCalc Software). Statistical analyses were conducted using IBM SPSS (version 22.0; SPSS Inc.). Data are expressed as mean (SD) for normally distributed continuous variables, median (25%-75%) for non-normally distributed continuous variables, and count SPI for categorical variables. *P* values <.05 were considered to be statistically significant.

Results

In total, 120 patients were enrolled, of whom 20 dropped out because of failure of PPG data storage (n=8), failure of SPI data storage (n=2), abnormal SPI data that could not be included in the analysis (n=9), and failure to measure VAS after surgery (n=1). Thus, 100 patients were included in the final analysis. The characteristics of these patients are summarized in Table 1. All but 1 patient had consciousness values of 2 points (conscious), as assessed by the modified Aldrete score at the time of PACU arrival. One patient scored 1 (arousable on

calling) who later had a value of 2 points upon leaving the PACU.

Individual changes in the spectrogram–CNN index and SPI without and with pain are presented in Figure 3. The 7 patients who had no postoperative pain were excluded from this analysis. In the case of pain, the mean spectrogram–CNN index and SPI values increased significantly (baseline spectrogram–CNN index: 28.5 (SD 22.1) versus PACU spectrogram–CNN index: 65.7 (SD 25.4), $P<.01$ in paired t test; baseline SPI: 42.5 (SD 16.7) versus PACU SPI: 53.5 (SD 17.8), $P<.01$ in paired t test). The area under the curve (AUC) of the ROC and cutoff values for detecting pain in terms of spectrogram–CNN index and SPI are listed in Table 2. The spectrogram–CNN index was statistically superior to the SPI (pairwise comparison of ROC curves: spectrogram–CNN index versus SPI, $P<.01$). Moreover, as shown in Table 2 and Figure 4, the spectrogram–CNN index showed improved performance measures in terms of balanced accuracy, sensitivity, and especially specificity.

Table 1. Characteristics of the study population.

Characteristic	Patients (N=100)
Male/Female	44/56
Age (years), mean (SD)	53.4 (12.5)
Height (cm), mean (SD)	161.8 (8.5)
Weight (kg), mean (SD)	62.2 (12.2)
ASA PS[a] 1/2/3	22/76/2
Type of surgery, n	
Breast	29
Colorectal	19
Hepatobiliary	12
Stomach	30
Thyroid	10
Postoperative pain intensity at PACU[b], n	
No (VAS[c]=0)	7
Mild (0 < VAS ≤ 30)	11
Moderate (30 < VAS ≤ 70)	58
Severe (70 < VAS ≤ 100)	24

[a]ASA PS: American Society of Anesthesiologists Physical Status.

[b]PACU: postanesthesia care unit.

[c]VAS: Visual Analog Scale (0=no pain; 100=the most severe pain).

Figure 3. Individual changes (n=100) in the spectrogram-convolutional neural network index (SCI, A) and the Surgical Pleth Index (SPI, B) and without and with pain. *$P<.05$ vs. baseline. The black circles represent the average of 5 min of the SCI or SPI observed without and with pain. The red circles indicate mean values for all patients. PACU: postanaesthesia care unit.

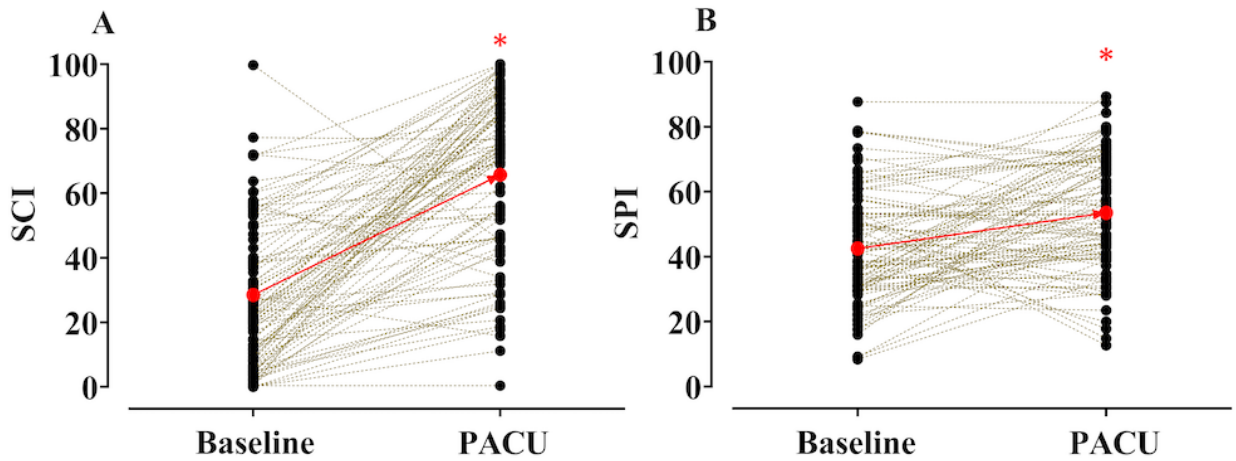


Table 2. Areas under the receiver-operating characteristic curves (AUCs) and cutoff values for assessing pain in the spectrogram-convolutional neural network index and surgical pleth index (SPI) in surgical patients.

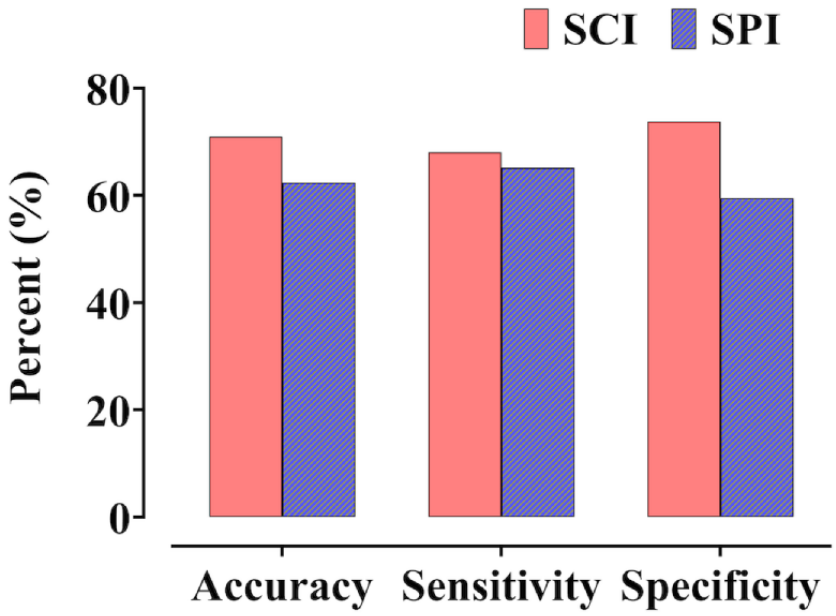
Parameter	Spectrogram–CNN[a] index			SPI
	Training set	Validation set	Test set	
AUC (95% CI)	0.992 (0.991-0.993)	0.932 (0.921-0.942)	0.757 (0.746-0.768)	0.659 (0.646-0.671)
P value	<.01	<.01	<.01	<.01
Cutoff value[b]	50	54	48	44
Sensitivity, %, specificity, %	94.6, 96.4	83.1, 88.2	68.1, 73.8	65.2, 59.5
Balanced accuracy, %[c]	95.5	85.7	71.0	62.4

[a]CNN: convolutional neural network.

[b]Cutoff values used for the calculation of sensitivity and specificity were calculated as “best fit” (highest combined sensitivity and specificity).

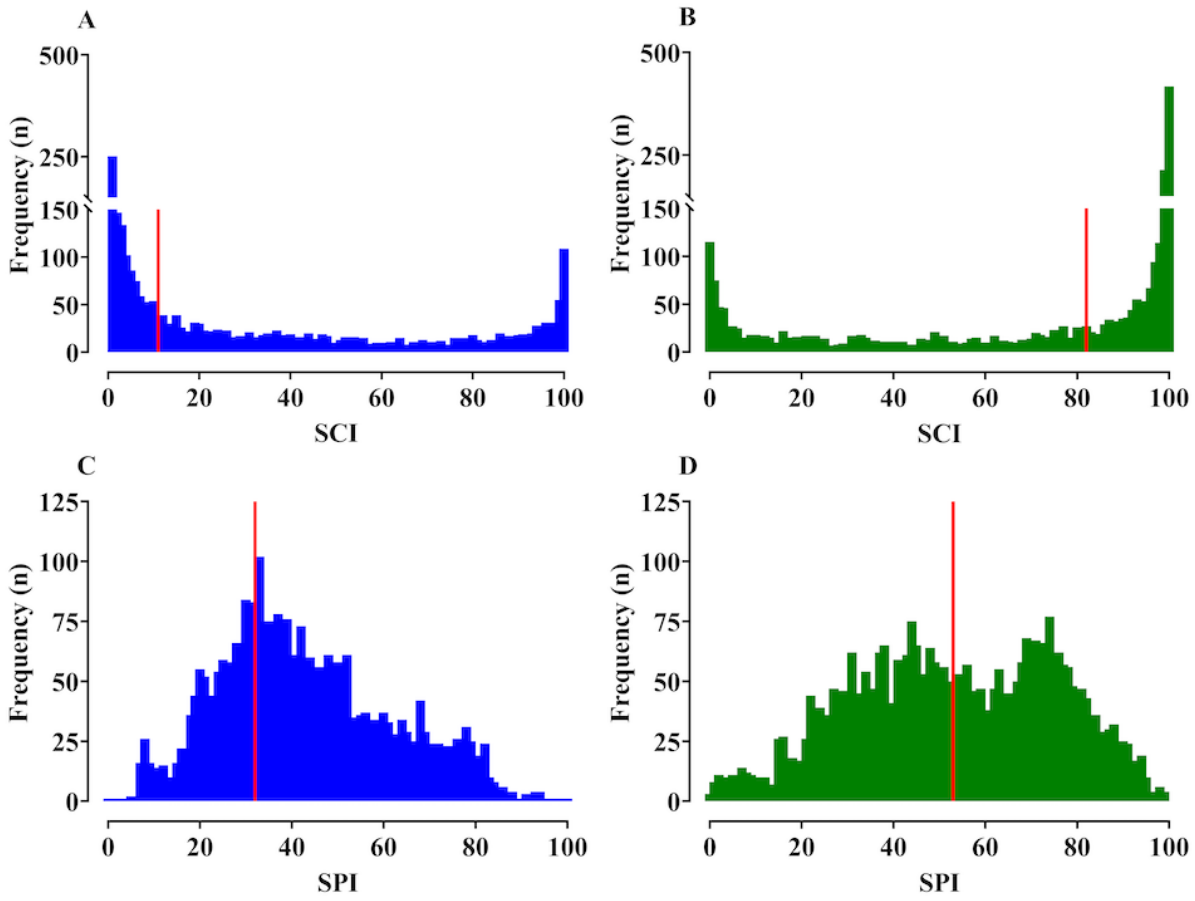
[c]Balanced accuracy is the corrected accuracy of the imbalance of a class set, calculated as (sensitivity + specificity)/2.

Figure 4. Performance measures of spectrogram-convolutional neural network index (SCI, red) and the Surgical Pleth Index (SPI, blue) in terms of accuracy, sensitivity and specificity. Accuracy means balanced accuracy. Cut-off value of SCI was 48.



Neither spectrogram–CNN index (mild: 69.5 [53.2-87.7], moderate: 71.8 [40.5-89.3], severe: 74.6 [62.9-84.7], *P*=.78, Kruskal–Wallis one-way ANOVA on ranks) nor SPI (mild: 53.5 [SD 21.0], moderate: 51.8 [SD 18.0], severe: 57.9 [SD 15.7], *P*=.37; one-way ANOVA) could statistically distinguish between mild, moderate, and severe pain. The frequency distributions of the spectrogram–CNN index and SPI values observed without and with pain during the data collection period are shown in Figure 5. The distribution of SPI values overlapped for with and without pain, suggesting that the SPI shows several false positives/false negatives, whereas the distribution of spectrogram–CNN index values showed a significant difference (*P*<.05) in patients with and without pain, suggesting that the spectrogram–CNN index can distinguish pain more clearly than the SPI.

Figure 5. Frequency distribution of the spectrogram-convolutional neural network index (SCI, A and B) and surgical pleth index (SPI, C and D) values observed without (A and C) and with (B and D) pain. During the data collection period (baseline: 5 min, postanaesthesia care unit: 5 min), the SCI and SPI values were observed every 10 sec. The vertical red lines show the median frequency (A: 11, B: 82, C: 32, D: 53).



Discussion

The spectrogram–CNN index proposed in this study outperformed the commercialized SPI pain index in the postoperative pain assessment of conscious patients. One of the main reasons for its outperformance is the spectrogram input containing the PPG's intact waveform information, making it possible to use hidden pain-related factors that could not be provided only by the peaks. Because existing pain assessment methods, such as the SPI, use only certain features of a PPG that show significant changes in surgical stimuli, numerous pain-related information reflected in the PPG may be overlooked. To overcome these limitations, research has been conducted on new pain-related features derived from sophisticated PPG waveform analysis in addition to the heart rate interval and PPG amplitude reflected in the SPI [5-7]. However, these pain indicators still depend on complex processes, such as peak detection and feature extraction, which require accurate peak extraction algorithms and are vulnerable to signal quality degradation. The proposed PPG spectrogram input does not require any peak detection or feature extraction process, thus avoiding problems such as peak misdetection during preprocessing. In addition, the spectrogram provides information from almost the whole PPG waveform because only the domain in which information is represented is transformed, while the underlying information is retained. This feature of the proposed model is its differentiating factor from SPI, which requires a complicated process of extracting pain features from biosignals, including PPG pulsation start and systolic maximum detection and verification. These preprocessing steps are necessary in existing pain assessment methods, but they are cumbersome and vulnerable, providing only limited features. Therefore, a simplified preprocessing process that still provides plentiful feature information can significantly improve the robustness of pain assessment.

Another key technique proposed in this study is the discrimination of postoperative pain using machine learning. Although machine learning–based pain assessment has already been studied [29-31], it is not suitable for practical clinical situations because it depends on high-dimensional clinical data, such as patient records and electroencephalograms, which are rarely used in postoperative care. However, machine learning from PPGs, which are frequently used in clinical practice during postoperative care, has high practical utility. In this study, we used a spectrogram–CNN combination, which has already been applied to electrocardiograms and electroencephalograms and has shown reasonable performance in predicting seizures and atrial fibrillation [32-35]. The spectrogram converts the data into 2D, and the CNN has the advantage of extracting the spatial features of multidimensional data. The combination of these

techniques thus extends the dimensions of PPG and allows spatiotemporal analysis, maximizing the use of features inherent in the signal. Nonlinear classification may be another important reason for the good performance of the proposed model. While the SPI is derived from a simple linear combination of normalized heartbeat interval and normalized pulse wave amplitude [3,36], the proposed spectrogram–CNN model performs nonlinear classification using the ReLU activation function. In addition, the nonlinearity is increased because the ReLUs are overlapped with each other in a multilayered structure.

In this study, patient data used for model development and validation were separated, and fivefold cross-validation was performed to eliminate interindividual interference and to generalize the model. Therefore, the proposed spectrogram–CNN index is expected to show similar performance in other groups of patients of similar age who did not participate in model development. However, there may be a few new variables to consider. In Table 2, performance measures were approximately 95%, approximately 85%, and approximately 70% for the training, validation, and test sets, respectively. The decision criteria also differed with 50 in the training set, 54 in the validation set, and 48 in the test set, indicating that overfitting occurred. In a random permutation test on balanced accuracy, the average accuracy was 49.5 (SD 0.7), and the range of values was 47.6-51.8. Considering that all balanced accuracies of the training, validation, and test sets were over 71, it is likely that the overfitting in our result stemmed from large intersubject variability rather than the model itself. Overfitting can be interpreted as degradation of the model's versatility, but it can also be interpreted as higher performance, at least in terms of validation accuracy, if sufficient data are used. Therefore, further studies are required to improve the reliability and versatility of this index using a large patient population with diverse body characteristics.

The spectrogram–CNN index may not have been able to distinguish pain intensity because the number of observations was relatively small. A previous study showed that SPI distinguishes postoperative pain intensity [28]; however, the authors of this previous study analyzed 1300 observations, whereas we only used 93. Nonetheless, a more fundamental reason is that the severity of pain was not accounted for when developing the spectrogram–CNN index. If sufficient PPG data were provided when learning to classify pain severity, it would be possible to classify pain intensity.

The database used in this study was shared by 2 research groups. Another group, independent of this study, has already published their results. Their pain classifier, based on a deep belief network using various PPG features, discriminated well between the presence and absence of pain [37]. Compared to the other study that extracted various features, ours used a simple spectrogram containing all PPG information without any complicated feature extraction procedure and is based on a CNN optimized for the spectrogram input. CNNs have lead in the machine learning field because they demonstrated performance improvements in image recognition in 2012 [9]. Moreover, the previous study evaluated pain based on full-length data, whereas ours assessed pain every 10 seconds, the same data display interval as in the SPI. Therefore, it can be applied to real-time pain assessment, which is in contrast to the other study.

There are some limitations to this study. First, the SPI was selected as a comparative index to evaluate the performance of the spectrogram–CNN index, but its suitability is somewhat debated. The SPI is neither developed nor recommended for use in conscious patients. However, when a new analgesic index is developed, it is essential to evaluate its performance, and this is commonly done by comparing it with an existing index using the same data. Although the SPI is not recommended for use in conscious patients, some studies on conscious patients suggest that SPI can discriminate between the presence and absence of pain [28,38]. Between the SPI and another commonly used commercial analgesic index—the Analgesia Nociception Index (PhysioDoloris, MetroDoloris)—the AUC–ROC for detecting postoperative pain in conscious patients was highest for the SPI [38]. Hence, the SPI was chosen as the comparative index for this study. Second, neostigmine and glycopyrrolate, when administered to reverse neuromuscular blockade, can contribute to PPG signals. As this study was observational, only data necessary for the development of a new analgesic index were collected during the normal anesthesia process without intervention. Neostigmine and glycopyrrolate were used in all patients because none of them required sugammadex. Neostigmine is known to be rapidly eliminated from the plasma after administration, with an average half-life of approximately 25 minutes [39]. We collected postoperative PPG data an average of 29.4 minutes after administration of these 2 agents. A previous study reported that baroreflex sensitivity was restored to its baseline value after approximately 82 minutes of glycopyrrolate administration [40]. It is possible that glycopyrrolate has mixed effects on postoperative PPG. However, because sugammadex usage is not common, it may be more beneficial to develop an index to distinguish pain based on data that can be obtained from actual practical conditions. Further studies are required to evaluate the extent of the effect of these 2 drugs on postoperative PPG. Third, it is difficult to determine whether the PPG data collected from the PACU solely reflect pain. In the conscious state, the sympathetic nervous system may be activated for other reasons, such as arousal or anxiety. Anxiety has been associated with reduced heart rate variability and vagal tone [41]. As we did not evaluate patient anxiety, we cannot determine its contribution to the PPG data. Considering the condition of patients who arrived in the PACU immediately after surgery, the PPG data mostly reflected immediate postoperative pain without controlling consciousness.

In conclusion, although there were several limitations to the study design, we confirmed that the newly developed spectrogram–CNN index can effectively detect postoperative pain in conscious patients. Further validation studies are required to assess its feasibility and prevent overfitting to various populations, including patients under general anesthesia.

Acknowledgments

This research was supported by two grants from the Basic Science Research Program through the National Research Foundation of Korea (NRF), one funded by the Ministry of Science and ICT (NRF-2018R1A4A1025704), and one funded by the Ministry of Education (NRF-2018R1D1A3B07046442), Republic of Korea.

Conflicts of Interest

None declared.

References

1. Chou R, Gordon DB, de LOA, Rosenberg JM, Bickler S, Brennan T, et al. Management of postoperative pain: a clinical practice guideline from the American Pain Society, the American Society of Regional Anesthesia and Pain Medicine, and the American Society of Anesthesiologists' Committee on Regional Anesthesia, Executive Committee, and Administrative Council. J Pain 2016 Feb;17(2):131-157. [doi: 10.1016/j.jpain.2015.12.008] [Medline: 26827847]
2. Logier R, Jeanne M, De Jonckheere J, Dassonneville A, Delecroix M, Tavernier B. PhysioDoloris: a monitoring device for analgesia / nociception balance evaluation using heart rate variability analysis. Annu Int Conf IEEE Eng Med Biol Soc 2010;2010:1194-1197. [doi: 10.1109/IEMBS.2010.5625971] [Medline: 21095676]
3. Huiku M, Uutela K, van Gils M, Korhonen I, Kymäläinen M, Meriläinen P, et al. Assessment of surgical stress during general anaesthesia. Br J Anaesth 2007 May;98(4):447-455 [FREE Full text] [doi: 10.1093/bja/aem004] [Medline: 17329347]
4. Ledowski T. Objective monitoring of nociception: a review of current commercial solutions. Br J Anaesth 2019 Aug;123(2):e312-e321 [FREE Full text] [doi: 10.1016/j.bja.2019.03.024] [Medline: 31047645]
5. Park C, Yang M, Choi B, Jeon B, Lee Y, Shin H, et al. Performance of the nasal photoplethysmographic index as an analgesic index during surgery under general anaesthesia. Sci Rep 2020 Apr 28;10(1):7130 [FREE Full text] [doi: 10.1038/s41598-020-64033-0] [Medline: 32346057]
6. Yang YL, Seok HS, Noh G, Choi B, Shin H. Postoperative Pain Assessment Indices Based on Photoplethysmography Waveform Analysis. Front Physiol 2018;9:1199 [FREE Full text] [doi: 10.3389/fphys.2018.01199] [Medline: 30210363]
7. Seok HS, Choi B, Noh G, Shin H. Postoperative Pain Assessment Model Based on Pulse Contour Characteristics Analysis. IEEE J Biomed Health Inform 2019 Nov;23(6):2317-2324. [doi: 10.1109/JBHI.2018.2890482] [Medline: 30605112]
8. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
9. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun. ACM 2017 Jun;60(6):84-90 [FREE Full text] [doi: 10.1145/3065386]
10. Pham H. Springer Handbook of Engineering Statistics. Cham, Switzerland: Springer Science & Business Media; 2006.
11. Schnider TW, Minto CF, Gambus PL, Andresen C, Goodale DB, Shafer SL, et al. The influence of method of administration and covariates on the pharmacokinetics of propofol in adult volunteers. Anesthesiology 1998 May;88(5):1170-1182 [FREE Full text] [doi: 10.1097/00000542-199805000-00006] [Medline: 9605675]
12. Minto CF, Schnider TW, Egan TD, Youngs E, Lemmens HJ, Gambus PL, et al. Influence of age and gender on the pharmacokinetics and pharmacodynamics of remifentanil. I. Model development. Anesthesiology 1997 Jan;86(1):10-23 [FREE Full text] [doi: 10.1097/00000542-199701000-00004] [Medline: 9009935]
13. Aldrete JA. Modifications to the postanesthesia score for use in ambulatory surgery. J Perianesth Nurs 1998 Jun;13(3):148-155. [doi: 10.1016/s1089-9472(98)80044-0] [Medline: 9801540]
14. Flanagan JL. Speech Analysis Synthesis and Perception. Cham, Switzerland: Springer Science & Business Media; 2013.
15. Wang W, Zhang G, Yang L, Balaji V, Elamaran V, Arunkumar N. Revisiting signal processing with spectrogram analysis on EEG, ECG and speech signals. Future Generation Computer Systems 2019 Sep;98:227-232. [doi: 10.1016/j.future.2018.12.060]
16. Huang J, Chen B, Yao B, He W. ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network. IEEE Access 2019;7:92871-92880. [doi: 10.1109/access.2019.2928017]
17. Alian AA, Shelley KH. Photoplethysmography. Best Pract Res Clin Anaesthesiol 2014 Dec;28(4):395-406. [doi: 10.1016/j.bpa.2014.08.006] [Medline: 25480769]
18. Shin H, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Trans Med Imaging 2016 May;35(5):1285-1298 [FREE Full text] [doi: 10.1109/TMI.2016.2528162] [Medline: 26886976]
19. Romero A, Gatta C, Camps-Valls G. Unsupervised Deep Feature Extraction for Remote Sensing Image Classification. IEEE Trans. Geosci. Remote Sensing 2016 Mar;54(3):1349-1362. [doi: 10.1109/tgrs.2015.2478379]
20. Soleymani S, Dabouei A, Kazemi H, Dawson J, Nasrabadi N. Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification. New York, NY: IEEE; 2018 Presented at: 24th International Conference on Pattern Recognition (ICPR); August 20-24, 2018; Beijing, China. [doi: 10.1109/icpr.2018.8545061]
21. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv 2021.

22. Raghu M, Poole B, Kleinberg J, Ganguli S, Dickstein J. On the expressive power of deep neural networks. In: Proceedings of the 34th International Conference on Machine Learning (Volume 70). 2017 Presented at: 34th International Conference on Machine Learning; August 2017; Sydney, Australia URL: https://dl.acm.org/doi/pdf/10.5555/3305890.3305975
23. Arora R, Basu A, Mianjy P, Mukherjee A. Understanding deep neural networks with rectified linear units. arXiv 2021.
24. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 2014;15(1):1929-1958 [FREE Full text]
25. Rubinstein RY, Kroese DP. The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning. Cham, Switzerland: Springer Science & Business Media; 2013.
26. Kingma DP, Ba JA. A method for stochastic optimization. arXiv 2021.
27. Bridle J. Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. Adv Neural Inform Process Syst 1990;2:211-217.
28. Thee C, Ilies C, Gruenewald M, Kleinschmidt A, Steinfath M, Bein B. Reliability of the surgical Pleth index for assessment of postoperative pain: a pilot study. Eur J Anaesthesiol 2015 Jan;32(1):44-48. [doi: 10.1097/EJA.0000000000000095] [Medline: 24827966]
29. Gram M, Erlenwein J, Petzke F, Falla D, Przemeck M, Emons MI, et al. Prediction of postoperative opioid analgesia using clinical-experimental parameters and electroencephalography. Eur J Pain 2017 Feb;21(2):264-277. [doi: 10.1002/ejp.921] [Medline: 27470494]
30. Lötsch J, Ultsch A, Kalso E. Prediction of persistent post-surgery pain by preoperative cold pain sensitivity: biomarker development with machine-learning-derived analysis. Br J Anaesth 2017 Oct 01;119(4):821-829 [FREE Full text] [doi: 10.1093/bja/aex236] [Medline: 29121286]
31. Tighe PJ, Harle CA, Hurley RW, Aytug H, Boezaart AP, Fillingim RB. Teaching a Machine to Feel Postoperative Pain: Combining High-Dimensional Clinical Data with Machine Learning Algorithms to Forecast Acute Postoperative Pain. Pain Med 2015 Jul;16(7):1386-1401 [FREE Full text] [doi: 10.1111/pme.12713] [Medline: 26031220]
32. Raghu S, Sriraam N, Temel Y, Rao SV, Kubben PL. EEG based multi-class seizure type classification using convolutional neural network and transfer learning. Neural Netw 2020 Apr;124:202-212. [doi: 10.1016/j.neunet.2020.01.017] [Medline: 32018158]
33. Zhai X, Jelfs B, Chan RHM, Tin C. Self-Recalibrating Surface EMG Pattern Recognition for Neuroprosthesis Control Based on Convolutional Neural Network. Front Neurosci 2017;11:379 [FREE Full text] [doi: 10.3389/fnins.2017.00379] [Medline: 28744189]
34. Parvaneh S, Rubin J, Rahman A, Conroy B, Babaeizadeh S. Analyzing single-lead short ECG recordings using dense convolutional neural networks and feature-based post-processing to detect atrial fibrillation. Physiol Meas 2018 Aug 23;39(8):084003. [doi: 10.1088/1361-6579/aad5bd] [Medline: 30044235]
35. Xia Y, Wulan N, Wang K, Zhang H. Detecting atrial fibrillation by deep convolutional neural networks. Comput Biol Med 2018 Feb 01;93:84-92. [doi: 10.1016/j.compbiomed.2017.12.007] [Medline: 29291535]
36. Ahonen J, Jokela R, Uutela K, Huiku M. Surgical stress index reflects surgical stress in gynaecological laparoscopic day-case surgery. Br J Anaesth 2007 Apr;98(4):456-461 [FREE Full text] [doi: 10.1093/bja/aem035] [Medline: 17350969]
37. Lim H, Kim B, Noh G, Yoo SK. A Deep Neural Network-Based Pain Classifier Using a Photoplethysmography Signal. Sensors (Basel) 2019 Jan 18;19(2):384 [FREE Full text] [doi: 10.3390/s19020384] [Medline: 30669327]
38. Lee J, Choi B, Jung Y, Lee Y, Bang J, Noh G. Evaluation of Surgical Pleth Index and Analgesia Nociception Index as surrogate pain measures in conscious postoperative patients: an observational study. J Clin Monit Comput 2020 Oct;34(5):1087-1093. [doi: 10.1007/s10877-019-00399-5] [Medline: 31628569]
39. Calvey TN, Wareing M, Williams NE, Chan K. Pharmacokinetics and pharmacological effects of neostigmine in man. Br J Clin Pharmacol 1979 Mar;7(2):149-155 [FREE Full text] [doi: 10.1111/j.1365-2125.1979.tb00915.x] [Medline: 216382]
40. Parlow JL, van Vlymen JM, Odell MJ. The duration of impairment of autonomic control after anticholinergic drug administration in humans. Anesth Analg 1997 Jan;84(1):155-159. [doi: 10.1097/00000539-199701000-00028] [Medline: 8989017]
41. Miu AC, Heilman RM, Miclea M. Reduced heart rate variability and vagal tone in anxiety: trait versus state, and the effects of autogenic training. Auton Neurosci 2009 Jan 28;145(1-2):99-103. [doi: 10.1016/j.autneu.2008.11.010] [Medline: 19059813]

Abbreviations

AUC: area under the curve
CNN: convolutional neural network
PPG: photoplethysmogram
ReLU: rectified linear unit
SPI: surgical pleth index
VAS: Visual Analog Scale

Edited by G Eysenbach; submitted 27.08.20; peer-reviewed by K Chen, H Demir; comments to author 21.09.20; revised version received 21.09.20; accepted 18.01.21; published 03.02.21.

Please cite as:
Choi BM, Yim JY, Shin H, Noh G
Novel Analgesic Index for Postoperative Pain Assessment Based on a Photoplethysmographic Spectrogram and Convolutional Neural Network: Observational Study
J Med Internet Res 2021;23(2):e23920
URL: http://www.jmir.org/2021/2/e23920/
doi:10.2196/23920
PMID:33533723

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org